CLUSTER ANALYSIS OF SOYBEAN PATHOGEN-RESPONSIVE GENES AND FUNCTIONAL CHARACTERIZATION IN ARABIDOPSIS

BY

YANYU HE

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Master's Committee:

Assistant Professor Steven J. Clough, Chair
Professor Lila O. Vodkin
Professor Ray Ming

Abstract

Defense against pathogens involves coordinated activation/deactivation of thousands of genes. When plants defend against some pathogens, the presence/absence of a single gene can make all the difference between resistance and susceptibility. For other diseases, like white mold caused by *Sclerotinia sclerotiorum* and sudden death syndrome (SDS) caused by *Fusarium virguliforme*, defense requires several genes, and therefore defense is controlled by quantitative trait loci (QTL). To understand plant defense and the genes involved, patterns of gene expression were used to identify genes that tend to be coordinately regulated across multiple disease reactions. Hierarchical clustering of soybean gene expression in response to pathogens including *S. sclerotiorum*, *F. virguliforme*, and *Pseudomonas syringae*, as well as soybean response to various non-pathogenic treatments, allowed for the identification of 11 candidate pathogen-specific responsive genes. Full-length cDNA of six candidate pathogen responsive genes of interest were cloned into *Escherichia coli*, and two of them, Glyma07g05480.1 (an O-methyltransferase) and Glyma18g45260.1 (a dihydroflavonol-4-reductase/cinnamoyl-CoA reductase), were cloned into an *Agrobacterium tumefaciens* binary vector and transformed into *Arabidopsis thaliana* to determine if these genes have a cross-species effect on enhancing disease resistance. This ongoing research project provides genes to be used for promoter analysis and identifies genes specific to pathogen infection. Genes will later be transformed into soybean to determine if they enhance resistance, and if so, their sequences can be developed into molecular markers to assist breeders in development of more resistant varieties and possibly used to develop transgenics with enhanced resistance.

To my family and friends

Acknowledgments

This project would not have been possible without the support of many people. First, I would like to express my appreciation to my advisor Dr. Steven Clough for his patience and thorough mentoring that helped me grow into a good researcher. I would like to extend my thanks to my graduate committee members Dr. Lila Vodkin and Dr. Ray Ming for their consideration and helpful feedback. Additionally, I owe a debt of gratitude to David Neece for his guidance of my laboratory work, which ensured my results were successful. I am also grateful to my colleagues Bernarda Calla, Osman Radwan, Doris Carbajulca, and as well as the undergraduates in the laboratory for their intellectual expertise and assistance with plant and laboratory materials for my research. And finally, I thank Charles Pignon and my wonderful friends, who believed in me and provided their care all along the way. And most importantly, I give a special thanks to my parents and family who are always there for me, offering their altruistic support and great love.

Table of Contents

CHAPTER I

GENERAL INTRODUCTION

I. OVERVIEW

This project was part of a multi-laboratory endeavor to identify pathogen-responsive genes in soybean and to characterize their promoters. My specific task was to identify and clone pathogen-inducible gene candidates, and to transform these candidate defense genes into *Arabidopsis thaliana* to determine if they have cross-species effects on defense to pathogens. Other labs in the project focused on the analysis of promoters of the pathogen-responsive genes that I identified.

II. LITERATURE REVIEW

High-throughput gene expression analysis methods

Soybean (*Glycine max*) is one of the most important crops in the structure of worldwide agriculture. Soybean disease caused by pathogens could devastate the crops and tremendously reduce crop production. Many observations on effects of disease have been reported from the studies conducted in major soybean growing areas in the United States (for example: Erwin and Ribeiro 1996; Hartman et al., 1991). The USA soybean industry is characterized by very large fields with little genetic diversity, making it particularly vulnerable to epidemic infection.

Disease control is a crucial component of row crop management. While fungicides may be an effective means of managing the risk of some disease, they increase production costs and can negatively impact the environment. In addition, improper management can lead to the prevalence of fungicide-resistant pathogens. A number of agronomical techniques are employed to reduce this risk, including crop rotation, soil management, and organization of the

plant growth season. Another powerful tool is the development of disease resistant varieties through both conventional breeding and marker assisted selection, which is a more sustainable approach and avoids fungicide risks.

Selection of disease resistance traits begins with the identification of resistance genes. These can often be found through traditional selective screening of recombinant populations, or by searching for genes up-regulated in response to pathogens. Improved resistance in cultivars is often associated with multiple pathogen-responsive genes, but few of them have been isolated and characterized, particularly these associated with partial resistance controlled by resistance quantitative trait loci (QTL). Their function and location in the genome are therefore often not precisely known. This gap of knowledge provides potential for furthering soybean disease resistance through precisely targeted gene selection and enhancement.

Many techniques have been used to measure the expression levels of specific genes or to characterize global expression profiles, such as northern blots, polymerase chain reaction after reverse transcription of RNA (RT-PCR), nuclease protection, cDNA sequencing, clone hybridization, expressed sequence tags (ESTs) and serial analysis of gene expression (SAGE). Among them, northern blots detect specific RNA species by electrophoresing the isolated RNA through agarose gel and separating the RNA species by size. Northern blotting offers a means to look closely at a particular gene expression pattern between tissues and pathogen infections. For better detection and quantification of gene expression from small amounts of RNA, transcript amplification boosts the development of PCR technologies. Compared with older methods used to measure mRNA abundance, differential display and northern blot, qRT-PCR is a low throughput technique with high specificity. Remarkable advances in sequencing

technology have made whole-genome and high-throughput expressed sequence tags (ESTs) sequencing practical and affordable and have paved the path to high-throughput gene expression analyses. The use of EST information allowed the development of SAGE and cDNA microarray, providing efficient means for rapid and high throughput quantification of a transcriptome. These modern methods have been put to use to identify the expression levels of thousands of genes at once to reveal transcription changes, and therefore, potentially important genes for further studies. To measure thousands of candidate target genes in a single experiment by microarray analysis provides a great chance to observe overall transcript patterns in cells and tissues. *Arabidopsis thaliana* response to pathogen was one of the first plant studies involving microarray technology (Scheideler et al. 2002). Infection with an avirulent *Pseudomonas syringae* triggered significant change in the expression of nearly 2000 genes within 7 hours. The encouraging results on Arabidopsis suggested that microarray technology could be used to study gene expression profiles in more important agricultural crops. Microarray platforms have been developed to support research in soybean (Vodkin et al., 2004) and numerous soybean-pathogen studies have been conducted (i.e.: Moy et al., 2004; Zou et al., 2005; Calla et al., 2009). Soybean microarray analysis is a powerful tool to assay transcriptional responses to infection in order to understand the fundamental mechanisms of plant defense against pathogens. The expression data used in these studies included the following pathogens: *P. syringae, Sclerotinia sclerotiorum, Fusarium virguliforme, Phytophthora sojae, Heterodera glycines, Phakopsora pachyrhizi,* and soybean mosaic virus. These pathogens limit soybean yield loss worldwide.

Soybean bacterial blight is a disease caused by *P. syringae*, a bacterium with a broad host range. Zou et al. (2005) studied gene expression patterns of soybean exposed to compatible and incompatible strains of this bacterium (*P. syringae* carrying or lacking the avirulence gene *avrB* was infiltrated into young unifoliates leaves), showing R-gene specific down-regulation of photosynthesis-related transcription during the incompatible interaction. In total, quantification of RNA levels from infected leaves showed that 3898 genes expressed significant changes in transcript abundance.

S. sclerotiorum is an important fungal pathogen that can infect a wide range of plant species. Soybean white mold is one disease that has caused tremendous damage to crop yields under disease-favorable weather conditions. A microarray screen was conducted on soybean stem tissue to analyze changes in gene expression between partially resistant and susceptible soybean genotypes at 8 and 14 hours post inoculation, and to identify genes involved in defense (Calla et al. 2009). 1270 significant genes from the comparison between time points and 105 genes from the comparison between genotypes were identified. PR-5 and anthocyanidin synthase genes were proved to be differentially expressed and located close to white mold resistance markers, making them good candidate white-mold-defense genes.

Sudden Death Syndrome (SDS) of soybean is a disease caused by the soil borne fungus *F. virguliforme*. The best way to control against this disease is through use of resistant germplasm, but in soybean resistance is only partial, with several poorly defined QTL providing some resistance. Radwan et al. (2011) used an Affymetrix microarray analysis to measure transcript abundance in resistant and susceptible roots at 5 and 7 days post infection, showing that 1694 genes expressed significant changes in response to *F. virguliforme*. The genotypes of 247 genes

were modulating in the resistant host, whereas those of 378 genes were modulating in the susceptible host. Also, comparison between locations of differentially expressed genes to known resistant QTL helped identify SDS defense-associated genes.

P. sojae is an aggressive soil borne pathogen that has the ability to infect soybean plants during several growth stages, especially early in the season when the weather is cooler and wet. It is the causal agent of soybean root rot, a prevalent disease in major soybean growing areas, resulting in devastating yield reduction each year. Moy et al. (2004) used soybean and pathogen mixed cDNA arrays of 4896 genes to observe effects that virulent *P. sojae* cast on soybean seedlings during a 48 hour time course, which showed that 1009 genes expressed significant changes. 820 genes were up-regulated, 517 of which were from the pathogen. Meanwhile the number of pathogen genes expressed peaked after 24 h, and the number of host genes expressed increased throughout the experiment (48 h).

H. glycines (*t*he soybean cyst nematode) is a plant-parasitic pest of soybean. It has become a major cause of crop loss not only in the US, but also around the world. Alkharouf et al. (2006) used microarrays of over 6,000 cDNA inserts to show specific changes in gene expression of susceptible soybean roots post inoculation by *H. glycines*. They identified genes, including WRKY6 transcription factor, lipoxygenase, phospholipase C, and chalcone reductase, which were differentially induced across most time-points. Several stress-related genes, phospholipase D and 12-oxophytodienoate reductase were induced during the early time-points. Finally, at 6 and 8 days post infection, there was an abundance of expressed transcripts encoding genes involved in transcription and protein synthesis.

Asian soybean rust (ASR) caused by *P. pachyrhizi* is a new serious threat to the world soybean industry. It was discovered for the first time in the continental US in November 2004 and spread from the eastern hemisphere. Panthee et al. (2007) used whole genome Affymetrix microarrays on *P. pachyrhizi*-exposed young soybean plants to perform transcriptome profiling of plants within 72 hours post inoculation, showing 112 genes differentially expressed after exposure to this pathogen, 46 of which were up-regulated. Most of the differentially expressed genes were general defense and stress-related genes.

Soybean mosaic virus (SMV) is a devastating virus disease of soybean that limits maximum yield potential. Bilgin et al. (2008) infected field-grown soybean plants with SMV while simultaneously fumigating them with ozone. Elevated O_3 treatment slowed systemic infection and disease development by inducing plant nonspecific resistance mechanisms. This inhibition of disease prevented the virus from impacting light-saturated photosynthesis. High-throughput gene expression analysis was then performed in a controlled environment. The transcripts of fungal, bacterial, and viral defense-related genes (such as PR-1, PR-5, PR-10 and EDS1) showed increase under elevated ozone.

Soybean microarrays have also been used to detail gene expression responses to the symbiotic microbe *Bradyrhizobium japonicum*, a species of legume-root nodulating, nitrogen-fixing bacteria. cDNA microarrays and Affymetrix chips were used to analyze gene expression in roots and isolated root hairs respectively at different time points post inoculation by *B. japonicum* (Brechenmacher et al., 2008; Libault et al., 2010). Several thousand genes were dramatically differentially expressed in roots in response to *B. japonicum* after 16 days of infection (Brechenmacher et al., 2008). Both microarray results showed reduced plant defense

response during nodule development, and a high level of regulatory complexity (transcriptional, post-transcriptional, translational, post-translational), likely essential for symbiosis development and adjustment to a nutritional balance richer in nitrogen.

Plant-microbe interactions are complex and dynamic, as both the pathogen and plants are constantly adjusting to different stages of infection and disease development. Therefore, making conclusions as to the physiological changes that might be occurring based on extrapolation of global expression data can be tricky. Alternatively, one could look at herbicides that might specially target a shared target of pathogens. An example would be the photosystem II (PSII) inhibiting herbicides, as some pathogen infections have also been shown to cause photo-inhibition. Therefore, herbicide studies were conducted using atrazine and bentazon, which inhibit the D1 subunit of PSII, and with glyphosate (Round-Up) which specifically inhibits biosynthesis of aromatic amino acids. Zhu et al. (2009) used microarray experiments to study transcription levels of approximately 36,000 genes in soybean leaf tissue treated with the PSII-inhibiting herbicides atrazine (lethal) and bentazon (non-lethal). Several hours after treatment, over 6,000 genes were differentially expressed, almost 90% of them showing similar expression patterns between both herbicides. The ability of soybean to metabolize bentazon allows it to survive exposure to this chemical. Plants exposed to bentazon quickly recovered, showing decreased amplitude of fold changes of most genes only 4 hours after treatment. As for the glyphosate, the majority of soybean grown in the US is resistant to it due to introduction of a bacterial enzyme that is insensitive to the chemical, and that biochemically conducts the same reaction as the plant-targeted enzyme. cDNA microarrays were used by Zhu et al. (2008) to study potential secondary effects of glyphosate. An increasing number of genes were expressed

up to 24 hours post exposure to glyphosate in sensitive plants; however, glyphosate-resistant plants showed only a slight genetic reaction one hour post treatment, followed by rapid adaptation to the herbicide. The result indicated that no major transcriptomic changes were associated with transgenic glyphosate-resistant soybean.

All of these studies identified soybean gene transcript responses to infection by pathogens, or to treatments that were related to disease. Global gene expression study by microarray analysis demonstrates a prospective for a better understanding of the difference between plant response to the same stimulus. Microarray analysis provides a means to assay transcriptional responses to plant stresses.

High-throughput DNA sequencing methods have recently led to the ability to sequence vast numbers of cDNA resulting in RNA-seq technology, providing a novel means to very precisely quantify a transcriptome (Wang et al., 2009). RNA-seq uses newly developed deep sequencing technologies and has clear advantages over other approaches. In contrast to microarray methods, sequence-based approaches directly determine the exact genes that are changing, whereas microarray results are always clouded with the possibility of unknown non-specific cross hybridization to a given DNA spot. RNA-seq carries such high expectation that it has started an evolution on eukaryotic transcriptome analysis method.

Clustering expression data to find genes of common expression patterns

Microarray output is often presented as a cluster matrix, with genes in rows and treatments as columns. Gene clustering is used to organize genes according to similarity in their expression patterns over several different treatments. Analysis of clustering patterns in genome-wide experiments can reveal various regulatory responses (Eisen et al., 1998). Because

genes with similar transcriptional behavior across samples may be involved in similar biological processes, clustering patterns can help identify genes with similar expression patterns (Quackenbush, 2001). In addition, because genes with similar functions are likely to be grouped together, clustering genes with a known function together with genes whose function is unknown can help give annotation to related unknown gene function.

Many statistical and computational methods have been developed for clustering. Hierarchical and K-means clustering are the two most widely used methods. Hierarchical clustering is best known for its application in sequence and phylogenetic analysis. When used for microarray analysis, a tree symbolizes inter-gene relationships, with branch lengths proportional to the degree of similarity between genes. While it is effective to organize genes in the original data table such that those with similar expression patterns are adjacent, this method cannot be used for large data sets because of its high computational intensity.

K-means clustering can be used to process datasets too large for hierarchical methods, but requires advanced knowledge about the number of clusters that the data should represent (Tavazoie et al., 1999). N genes are partitioned into K clusters, with K pre-determined by the user, with the following method: K arbitrary centroids are picked, and each gene is assigned to its closest centroid. The centroids are then adjusted to be the means of the genes assigned to them. Each gene is then reassigned to its closest centroid and each centroid readjusted to be the mean of its assigned genes. This process is repeated until no further changes are obtained. While this repetitive process can process large datasets, it depends on the use of an optimal K value and the program must be run with several values of K to determine an optimal one.

Both hierarchical and K-means clustering assume that each gene can be assigned only to one cluster. However, biologically, many genes may be involved in several physiological networks and controlled by various regulatory mechanisms. A fuzzy algorithm is therefore preferable for gene expression data analysis. Fuzzy K-means clustering is a robust and accurate method that was used for exploration of conditional co-regulation of yeast gene expression (Gasch and Eisen, 2002). This algorithm demonstrated its advantages in identifying key genes in microarray studies. This method provides 'one-to-many' mapping, which means that a single gene can be assigned to different clusters with a calculated degree of membership, which gives access to more sophisticated relations between a data object and its related clusters. This is a more suitable option for analysis of biological datasets as most genes are involved in various genetic networks and are governed by several regulatory mechanisms (Futschik and Kasabov, 2002).

Promoter analysis

A promoter is the region of gene that regulates transcription and determines the intensity and location of gene expression. A promoter is composed of different regulatory elements that control its function and specific induction. Analysis based on stably transformed tissues or transient expression analyses of promoters can reveal their strength and regulation in different tissues. Reproducible, rapid and quantifiable promoter analyses can be simplified with transient expression, via direct DNA introduction into protoplasts using electroporation (Christensen et al. 1992), or particle bombardment-mediated transformation into intact plant tissues (Rolfe and Tobin, 1991).

Pathogen resistance can occur through modification to a gene coding region or by altering gene expression patterns through changes to the promoter region. The former has been more extensively researched because modifications to the coding region are overall easier to analyze than promoters. In this project there was more focus on the promoter region of genes involved in pathogen resistance, as the expression pattern (location, timing and intensity of gene expression), which is regulated by promoters, is likely as important for pathogen response as the coding region. This part of the research project was the focus of our collaborating partners.

Multiple promoters can be acted up by the same transcription factor due to shared common sequences between regulatory regions of difference genes. WRKY proteins are an important family of plant transcription factors involved in pathogen defense regulation. The general binding preference of WRKY proteins suggest genes containing promoter elements (e.g. WRKY genes and other pathogen related defense genes) are likely to be the targets of WRKY factors (Rushton and Somssich, 1998). Many WRKY proteins are thought to regulate the response to biotic and abiotic stresses. Comparative expression studies with several Arabidopsis *WRKY* genes indicate that some of these family members are involved in regulating pathogen infection response. For example, in transgenic Arabidopsis plants, an *AtWRKY6* promoter–GUS reporter gene is strongly activated during response to infection by pathogenic bacteria (Quirino et al., 1999).

Arabidopsis transformation

Today, many agronomical and horticultural species of importance are routinely transformed using *Agrobacterium tumefaciens* or *Agrobacterium rhizogines*, and the list of

species that is susceptible to Agrobacterium-mediated transformation seems to continually grow.

Plant transformation is the introduction of a foreign gene into a plant genome, using either ballistics or *Agrobacterium* species. *A. tumefaciens* is a soil-borne bacterium that infects host plants through wounds in the crown region of a plant, and requires the presence of a tumor-inducing (Ti) plasmid which contains genes required for insertion of a segment of bacterial DNA (the T-DNA region) into the host's genome. By inserting the T-DNA into a plant's nuclear genome, the plant cells begin to produce large amounts of auxin and cytokinin in a manner that leads to disruption of normal cell growth and the formation of large tumors or galls. Researchers utilized the ability of Agrobacterium to transform plants as a valuable tool by replacing the tumor-inducing genes of the T-DNA with genes of interest, such as a selectable marker together with a defense-enhancing gene. Therefore, the altered *Agrobacterium* serves as a vector for plant transformation. Gene of interest, selectable marker (i.e. resistance to an antibiotic) and reporter gene (e.g. GUS, GFP) can be introduced into the plants. When performing molecular manipulations and plant transformation, the most convenient vector system is the binary vector system in which cis-acting elements required for plant transformation are contained within the vector (Hoekema et al., 1983). Binary tumor-inducing (Ti) plasmid vector systems were used to transform several plants (including *A. thaliana*) with a kanamycin resistance marker (An et al., 1986).

Once a gene in a plant like soybean is prioritized as having a certain role in a given physiology, such as disease resistance, that gene can be cloned and analyzed in other plants, such as a susceptible soybean genotype in the case of disease resistance, to verify its function.

An alternative to cloning a crop gene and expressing it in that species (which is often time consuming and technically challenging) is to clone the gene and overexpress it in Arabidopsis. For instance, a cloned watermelon transcription factor cDNA which was believed to be involved in defense to disease, *WRKY70*, functioned when cloned into Arabidopsis, providing the Arabidopsis with enhanced resistance to the fungal pathogen *Botrytis cinerea* and the bacterial pathogen *Erwinia carotovora* (Cho et al. 2012). The successful experiment suggests that full-length cDNA of soybean pathogen responsive genes could be cloned into an Agrobacterium binary vector and transformed into Arabidopsis to test if the genes of interest have an effect on enhancing disease resistance.

Expressing a crop gene in Arabidopsis has many benefits. Arabidopsis is easily transformable with *Agrobacterium* T-DNA (Clough and Bent, 1998). Plants and *Agrobacterium* are grown, and then floral dip is performed: the flowering plants are dipped in the suspended bacteria, and then grown until their seeds can be harvested. The collected seeds are grown on the selection medium, which kill all of the seeds except the transformants of interest. Additionally, one could also obtain Arabidopsis mutants of genes of interest. Several non-profit labs have generated and made readily available, a vast collection of T-DNA insertion mutants, which provides a high probability of finding an Arabidopsis mutant with a T-DNA inserted into one's gene of interest, with search and ordering easily done on-line from The Arabidopsis Information Resource website (http://www.arabidopsis.org/). Therefore, if one finds a gene in soybean, such as an *MMP2* gene, one could obtain Arabidopsis *MMP2* mutants and test for functional complementation of this mutant with the soybean gene, and thus provide evidence of a function of the soybean gene. Such complementary gene function between plant species

often works if the genes are well conserved. For example, using an ndr1 Arabidopsis mutant, Cacas et al. (2011) showed that a coffee NDR1 homolog functioned similarly as the well-characterized NDR1, as the coffee gene fully restored *P. syringae* resistance to this Arabidopsis mutant. Soybean is no different in regard to having some of its genes showing similar function in Arabidopsis. When a soybean PHD transcription factor was expressed in Arabidopsis, this soybean gene provided the transgenic Arabidopsis with enhanced resistance to salt stress (Wei et al., 2009).

III. OBJECTIVES

A. Search for soybean pathogen-responsive genes and characterize them

1. Cluster RNA expression data from microarray and qRT-PCR studies of soybean response to various pathogens.

2. Reduce the list of candidate pathogen-responsive genes by selecting genes that show little response to other stimuli.

B. Clone select soybean pathogen-responsive genes for over-expression in *Arabidopsis thaliana*

1. Use RT-PCR to obtain full-length cDNA of at least one of the current set of 11 pathogen-specific gene candidates.

2. Clone cDNA into an over-expression vector for plant transformation via Agrobacterium.

C. Inoculate the Arabidopsis plants that are over-expressing a soybean pathogen-responsive gene with *P. syringae* and determine rate of bacterial replication in mutant versus parent.

CHAPTER II

IDENTIFYING AND CHARACTERIZING SOYBEAN PATHOGEN-RESPONSIVE GENES

I. ABSTRACT

Plants cannot run from pathogen attacks, and therefore they must be able to sense life-threatening attacks and respond appropriately. The main goal of all pathogens is to acquire the nutrients and molecular building blocks that they need to grow and propagate. Therefore, because pathogens have many overlapping needs, it is not surprising that many defense responses are shared, independent of the type of pathogen, in addition to some defense responses being specific to individual pathogens. Soybean (*Glycine max*) high-throughput gene expression data from many disease studies was gathered to run a meta analysis totaling 5,417 genes and 54 treatments, including soybean response to the pathogens *Sclerotinia sclerotiorum*, *Fusarium virguliforme*, and *Pseudomonas syringae*. In addition, soybean transcription responses to various non-pathogenic treatments were also included. Hierarchical clustering and fuzzy k-means clustering of the data, followed by selective filtering for patterns of interest, allowed for the identification of 11 candidate pathogen-specific responsive genes.

II. INTRODUCTION

Plants show an active response to pathogen attacks, therefore defense reactions most likely involve changes of gene expression level and this has been proven in many studies. At least for some bacterial pathogens, the difference between effective and ineffective resistance largely depends on how robustly and rapidly the plant can orchestrate the degree of transcriptional changes of thousands of genes (Tao et al., 2003; Zou et al., 2005). The development of high-throughput gene expression technologies has led to the generation of vast

amounts of data on transcriptome responses to various treatments. Because gene expression profiles help identify mRNA species present at statistically different levels, they are likely to indicate which genes are involved in defense or susceptibility to pathogens, as well as which genes might be key to differentiating resistance and susceptibility.

Nowadays, high-throughput gene expression studies of plant–pathogen interactions are being applied more and more frequently. While other transcript profiling methods exist (e.g. differential display and serial analysis of gene expression), during the last decade cDNA microarray technology has been the most popular means of studying the responses of thousands of genes very quickly, allowing comprehensive sampling capacity. Today, with decreasing costs of RNA-seq, this amazing technology is expected to rapidly dominate the field and become the method of choice. Because of the thoroughness and power of microarray analysis, many studies have been published of gene expression changes in plants during pathogen infection and disease development. In soybean (*Glycine max*) this technology has been performed to obtain gene expression profiling in various pathogenic treatments (i.e.: Zou et al., 2005; Calla et al., 2009). Additionally, studies conducted previously in our laboratory demonstrated the soybean gene expression changes during the symbiotic response of nodule initiation (Brechenmacher et al., 2008), as well as leaf response to herbicide treatment (Zhu et al., 2009). These multiple high-quality datasets provide a rich source of gene expression for data mining genes involved in defense.

Mathematical clusters are convenient and powerful tools to identify genes of interest from gene expression studies. Cluster analysis for genome-wide expression data from DNA microarray hybridization uses statistical algorithms to arrange genes based on their similarities

in gene expression pattern, which are more likely to be functionally related (Quackenbush, 2001). Producing graphical results like heat maps facilitates identification of clear expression patterns of interest, such as identifying genes with common differential expression across multiple disease and stress conditions.

Clustering methods have been widely applied in modern research, such as hierarchical clustering and fuzzy k-means clustering. Hierarchical clustering method groups genes on the basis of similarity in the pattern among all samples. However, hierarchical clustering assumes that each gene can only be assigned to one cluster. This feature would limit the search for genes that participate in different genetic networks and governed by a variety of regulatory mechanisms. Therefore, fuzzy K-means clustering was developed by Gash and Eisen (2002) to address this facet of biology. This fuzzy algorism allows a single gene to be assigned to different clusters with calculated degrees of memberships, which displays a more complete relationship between a target gene and its related clusters.

The Soybean Gene Expression Database (SGED: http://sged.cropsci.illinois.edu/) was developed with in-house microarray expression. This web-based microarray database provides assistance to data management and interpretation and comes with several interfaces and applications that help explore and analyze SGED data. Tools include algorithms of K-means clustering, hierarchical clustering and fuzzy K-means, all with graphical outputs to help visualize gene expression patterns and identify co-expressed genes. SGED is a useful analysis tool and is highly integrated with the Soybean Gene Information Database (SGID), which provides supportive annotation and other background information about spots on soybean cDNA

microarrays and the soybean Affymetrix chip, as well as annotations related to Glyma IDs from Phytozome (http://www.phytozome.net/).

Comparing expression results from different treatments can reveal common or unique pathogen response pathways triggered upon exposure to pathogens, and aid in finding common promoter elements. For the analysis presented in this chapter, we used expression data from 17 experiments covering 54 treatments/timepoints (Table 2.1). Data was compared from four experiments on *Pseudomonas syringae* induced compatible on incompatible interactions several hours after inoculation, two experiments on soybean response to *Fusarium virguliforme* toxin (SDS disease) in roots and leaves 5 and 7 days after inoculation, four experiments on stem and leaf responses to *Sclerotinia sclerotiorum* in resistant or susceptible material, and responses of soybean to the nitrogen-fixing symbiont *Bradyrhizobium japonicum* from inoculated root hairs and nodulating whole roots. The meta analysis also included soybean response to purified chemical treatments: the herbicides glyphosate (inhibitor of aromatic amino acids) and bentazon (inhibitor of photosystem II, but degradable by soybean, thus causing a transient effect only) and oxalic acid (virulence factor released by *S. sclerotiorum*). Two final control treatments added to this thorough analysis were the effect of growing soybean roots in tissue culture ('hairy root' versus normal roots) and the effect of vacuum infiltration of 10 mM $MgCl_2$ (the control for the *P. syringae* inoculations).

III. METHODS

Data collection

Genes of interest were gathered into a pool from three sources: published papers, keyword searches in the SGED database and recent Clough laboratory results. Genes were also

selected from published papers where the genes were qRT-PCR verified: soybean unifoliate leaf response to *P. syringae* (Zou et al., 2005), stem response to *S. sclerotiorum* (Calla et al., 2009), seedling response to *P. sojae* (Moy et al., 2004), root response to *Heterodera glycines* (Alkharouf et al., 2006; Vaghchhipawala et al., 2001), leaf response to *Phakopsora pachyrhizi* (Panthee et al., 2007; van de Mortel et al., 2007; Choi et al., 2008), leaf response to soybean mosaic virus (Bilgin et al., 2008; Yang et al., 2011), and a study that looked at *R*-gene specific expression in response to pathogen (Graham et al., 2002).

To conduct the keyword searches, the in-house database, SGED was utilized. The terms that were used to search for genes of interest were: (1) AVR; (2) beta-1,3 glucanase; (3) chitinase; (4) elicitor; (5) GST; (6) hypersensitive; (7) pathogen; (8) disease; (9) resistance; (10) WRKY.

Finally, statistically significant gene lists from in-house microarray experiments of soybean inoculated with *P. syringae, S. sclerotiorum, F. virguliforme* was applied as well as soybean treated with different herbicides. Collecting all these genes and removing duplication, resulted in an initial list of 5,417 putative pathogen-responsive genes.

Treatments that were applied for cross comparison

The genes were clustered based on their expression patterns across 54 treatments (Table 2.1 is a summary of treatments used for this analysis). The treatments can be divided into four main categories: pathogenic, symbiotic, chemical and the other. There were 15 subsets describing each treatment, which divided each treatment into different time-points.

Hierarchical cluster of candidate genes

Hierarchical clustering in SGED performs clustering on both experiments and genes simultaneously, which provides a better understanding on the behavior of soybean pathogen-responsive genes under pathogen and non-pathogen treatments. A list of candidate genes was uploaded to the server and 54 treatments from multiple projects were selected. An image integrating a tree and a heatmap was produced as the hierarchical clustering result. A red color indicated gene induction in response to the conditions they were exposed to, green indicated gene repression under those conditions, and black indicated the absence of expression level change. Genes whose values were missing in one or more clustering experiments are ignored by the program. Furthermore, annotation for each gene can be retrieved in Excel files. The genes in the final dataset were classified according to their similarity in expression patterns.

Fuzzy K-means cluster of candidate genes

The clustering analysis presented here made use of the fuzzy K-means cluster method in the data of SGED. A specific number k (equals to twice the number of selected treatments) of initial centroids was applied on the sets entered. In order to visualize the result, each row represented the relative transcript abundance of a single gene while each column represented the relative transcript abundance of several genes, as measured for each hybridization. All genes in the dataset were related to each centroid with a membership score. Users can input a membership cutoff and cluster number to expand a specific gene cluster showing the expression pattern of interest; a cut-off value was set at 0.2 for a gene to be considered into a certain centroid. Centroids with less than 10 genes meeting the cut-off value of 0.2 were discarded. Only centroids of interest were selected for discussion.

IV. RESULTS

Microarray and qRT-PCR data was gathered and clustered across 54 treatments using hierarchical clustering. Values used for clustering were the log2 ratios of a treatment versus control, or of the expression change across time. Of the 5,417 pathogen-responsive genes analyzed, a cluster of 411 (Figure 2.1, Table 2.1) was identified as being coordinately induced in response to the pathogen treatments (*P. syringae, S. sclerotiorum* and *F. virguliforme*). A fuzzy K-means algorithm was used to analyze these 411 genes. Fuzzy K-means clustering output of expression data was then displayed as an overview on multiple centroids. Each row in the diagram represented one centroid categorized by statistical analysis. The Centroid 3 (Figure 2.2) was obtained based on the pattern of having genes that were up-regulated in pathogen experiments and an inversed or no expression change displayed in the nodulation studies, and weak or nonexistent change upon treatment with herbicides. Genes were assigned to this centroid using a membership cutoff at 0.2, producing a refined list of 56 candidate pathogen-inducible genes.

The 56 genes selected in Centroid 3 were hierarchically clustered and displayed to enhance visualization of expression patterns. To facilitate down-stream research projects that will look at gene expression in transgenic hairy-root cultures, the candidate list was reduced to 51 genes (Figure 2.3) by removing genes whose expression increased in hairy roots when compared to normal roots. Candidate genes are therefore expected to indicate no detectable expression changes in disease assays in hairy roots, unless it is due to a response to the particular pathogen studied.

In the list of 51 pathogen-responsive genes, it was noted that many were also induced in leaves by the stress of vacuum infiltrating 10 mM $MgCl_2$ into their apoplastic space (this treatment was a control in the *P. syringae* study). Therefore, using a fold-change cut-off value of about 1.5X (lg2 > 0.5), the list of 51 was reduced to 16 (Figure 2.4), which was further reduced to 11 genes (Figure 2.5) by removing five genes (Glyma02g18380.3, Glyma03g37650.1, Glyma09g03490.3, Glyma18g41820.1, Glyma18g44630.1) that showed very weak expression across all treatments. These 11 genes were chosen as the best pathogen-responsive candidates as their expression was clearly induced by pathogens, but not by the symbiotic bacterium *B. japonicum*, $MgCl_2$ vacuum infiltration, or in hairy roots. In addition, we added one more treatment for the final clustering, the expression of the gene during a root growth experiment to verify that the selected genes are not strongly changing during normal root growth. The 11 genes (Table 2.3) code for a variety of proteins including known defense genes (dirigent-like protein and two PR10/allergen Bet v I family) as well as two transporters (an ABC transporter and an ammonium transporter), a subunit of NADPH oxidase with homology to an iron reductase, a phospholipase, genes most likely involved in secondary metabolism (an aldo/keto reductase, a flavonol reductase/cinnamoyl-CoA-reductase, and an O-methyltransferase) and a defense-associated class of transcription factors (WRKY).

V. DISCUSSION

DNA microarrays make it possible to analyze the expression levels of thousands of genes simultaneously and they have been used extensively to deliver a perception of gene behavior in response to diverse pathogen attacks. For this project, we utilized the in-house SGED database clustering tools and soybean microarray datasets to conduct a meta analysis of soybean gene

expression in response to pathogens and non-pathogenic treatments to identify the most specific pathogen-responsive genes.

Our genes of interest were those with a pathogen-specific response. To help identify these best gene candidates, and refine the search, multiple non-pathogenic treatments were included in the clustering. Soybean response to the symbiotic *B. japonicum* will require the downregulation of expression of defense genes (Brechenmacher et al., 2008), and therefore genes were selected that were induced by multiple pathogens, but reduced or unaffected in two nodulation studies. Additionally, genes induced by the stress of vacuum infiltration or by hairy root induction were selected against, ensuring that the genes selected were fairly pathogen specific. Genes nonresponsive to herbicides were more likely pathogen-specific, although there is some possibility for treatment interaction. For example, a gene could be sensitive to inhibition of photosynthesis, and also expressed during pathogen attack as the pathogen may also target inhibition of photosynthesis as a virulence strategy.

As one of the ultimate goals of this meta analysis was to identify candidate defense-responsive promoters, a collaborating laboratory (John Finer, the Ohio State University) will test these promoters with promoter::gfp fusions in hairy root assays. Therefore, it was imperative that the genes identified not be affected simply by induction and propagation of hairy-root cultures, or by root growth, and therefore gene expression related to these effects were included.

Of the 11 identified pathogen-responsive genes, several look very interesting in terms of plant defense to pathogens. As transcription responses require transcription factors, it was ensuring to see that a WRKY factor (Glyma11g29720.1) was one of the 11 genes selected. WRKY

factors are the most important class of transcription factors in defense, as they appear to be strongly associated with biotic stress responses (reviewed in Ülker and Somssich, 2004). Another gene of interest is a dirigent (Glyma01g31750.1), as these genes are often induced by pathogens (Zou et al., 2005) and are involved in the synthesis of lignans and lignins (Davin and Lewis, 2000), both of which are often effective defense responses. Pathogenicity-related (PR) proteins are a hallmark of defense responses in plants (Van Loon et al., 2006) and therefore it was not surprising to see two genes that are located next to each other (Glyma17g03340.1 and Glyma17g03350.2) encode for PR10, often annotated a member of the allergy-inducing peptide family Bet v I and believed to be a ribonuclease (Van Loon et al., 2006).

One of the first responses to some pathogens is an oxidative burst that serves to prime and induce other defenses. A membrane-bound NADPH oxidase has been shown to be required for this oxidative burst to occur (Torres and Dangl, 2005), and therefore it was exciting to see that one of the pathogen-specific genes identified here, Glyma19g42220.1, encodes for a ferric reductase subunit of an NADPH oxidase complex. In addition to reactive oxygen species serving as defense signals, some signals are related phospholipid metabolism (Kachroo and Kachroo, 2009), and therefore the gene Glyma10g42820.3 is promising as it encodes a putative phospholipase.

Movement of nutrients and defense metabolites during infection require the presence and function of numerous transport proteins. Therefore, it was not surprising to see two transporters make the list of 11 pathogen-responsive genes. One, an ATP Binding Cassette (ABC) transporter (Glyma10g02370.2), might be involved in movement of toxins, and the other, Glyma18g43540.1 is homologous to ammonium transporters.

Secondary metabolites play a very important role in plant defense, and three genes that likely are involved in secondary metabolism were pathogen-specific. The phenylpropanoid pathway seems to always be induced in soybean in response to pathogens (Zou et al., 2005; Calla et al., 2009; Radwan et al., 2012) and one identified gene (Glyma18g45260.1) belongs to this pathway, having homology to both a dihydroflavonol-4-reductase and a cinnamoyl-CoA reductase. The other two genes (Glyma14g00870.1 and Glyma07g05480.1) encode for enzymes that could alter a substrate to detoxify or to perhaps lead it to being antimicrobial, one being an aldo/keto reductase, and the other an O-methyltransferase.

As the number of available datasets increases, these studies will be repeated to further refine the selection of pathogen-specific gene expression. Additionally, meta analysis will allow characterization of genes whose functions are currently unknown based on co-expression patterns across multiple treatments. Using clustering techniques to perform genome-wide expression analyses allows the comparison of different diseases and treatments and makes it possible to group genes of similar behavior. And therefore, genes of unknown function or that do not correspond to any known sequence can be assigned to defense highlighting another example of clustering methods as an invaluable tool for genetic expression analysis.

Tables and Figures

Table 2.1. Description of the experiments used in cluster analysis.

Label	Description	Reference
Nodulation		
Nod2T4dpi	Nodulation whole roots 4 days after infiltration	Brechenmacher et al. 2008
Nod2T8dpi	Nodulation whole roots 8 days after infiltration	Brechenmacher et al. 2008
Nod2T16dpi	Nodulation whole roots 16 days after infiltration	Brechenmacher et al. 2008
T6	Nodulation root hairs 6 hours	Brechenmacher et al. 2008
T12	Nodulation root hairs 12 hours	Brechenmacher et al. 2008
T18	Nodulation root hairs 18 hours	Brechenmacher et al. 2008
Pseudomonas syringae		
HRT2vsMgCl2T2	HR inducing strain infiltration vs 10mM MgCl2 control, 2hpi	Zou et al. 2005
VIRT2vsMgCl2T2	VIR inducing strain infiltration vs 10mM MgCl2 control, 2hpi	Zou et al. 2005
HRT8vsMgCl2T8	HR inducing strain infiltration vs 10mM MgCl2 control, 8hpi	Zou et al. 2005
VIRT8vsMgCl2T8	VIR inducing strain infiltration vs 10mM MgCl2 control, 8hpi	Zou et al. 2005
HRT24vsMgCl2T24	HR inducing strain infiltration vs 10mM MgCl2 control, 24hpi	Zou et al. 2005
VIRT24vsMgCl2T24	VIR inducing strain infiltration vs 10mM MgCl2 control, 24hpi	Zou et al. 2005
MgCl2T2vsT0	10mM MgCl2 control, 2hpi vs Null	Zou et al. 2005
AvrvMock2	Soybean HR williams82 (rpg1) with AvrB 2hpi	In-house
AvrvMock4	Soybean HR williams82 (rpg1) with AvrB 4hpi	In-house

Table 2.1. (cont.)

AvrvMock8	Soybean HR williams82 (rpg1) with AvrB 8hpi	In-house
rpg1_AvrB_v_Mock_2h	Soybean VIR Flambeau (rpg1) with AvrB 2hpi	In-house
rpg1_AvrB_v_Mock_4h	Soybean VIR Flambeau (rpg1) with AvrB 4hpi	In-house
rpg1_AvrB_v_Mock_8h	Soybean VIR Flambeau (rpg1) with AvrB 8hpi	In-house
Fusarium virguliforme		
R_Infect_v_Mock_5d_Root	SDS roots resistance infected vs mock 5 days after infiltration	Radwan et al. 2011
R_Infect_v_Mock_7d_Root	SDS roots resistance infected vs mock 7 days after infiltration	Radwan et al. 2011
S_Infect_v_Mock_5d_Root	SDS roots susceptible infected vs mock 5 days after infiltration	Radwan et al. 2011
S_Infect_v_Mock_7d_Root	SDS roots susceptible infected vs mock 7 days after infiltration	Radwan et al. 2011
S5_inf_v_mock_Leaves	SDS leaves susceptible infected vs mock 5 days after infiltration	In-house
S7_inf_v_mock_Leaves	SDS leaves susceptible infected vs mock 7 days after infiltration	In-house
R5_inf_v_mock_Leaves	SDS leaves resistance infected vs mock 5 days after infiltration	In-house
R7_inf_v_mock_Leaves	SDS leaves resistance infected vs mock 7 days after infiltration	In-house
Sclerotinia sclerotiorum		
T14vsT8_W	White mold stem 14hpi vs 8 hpi, Williams82	Calla et al., 2009
T14vsT8_PI	White mold stem 14hpi vs 8 hpi, PI194.639	Calla et al., 2009
T12T0_AC	White mold 12hpi vs T0, AC Colibri	In-house
T12T0_OXO	White mold 12hpi vs T0, OXO transgenic	In-house
T24T12_AC	White mold 24hpi vs T0, AC Colibri	In-house

Table 2.1. (cont.)

T24T12_OXO	White mold 24hpi vs T0, OXO transgenic	In-house
T24T0_AC	White mold 24hpi vs T0, AC Colibri	In-house
T24T0_OXO	White mold 24hpi vs T0, OXO transgenic	In-house
AC_12h_v_0h	AC Colibri 12hpi vs 0hpi	In-house
AC_24h_v_0h	AC Colibri 24hpi vs 0hpi	In-house
AC_36h_v_0h	AC Colibri 36hpi vs 0hpi	In-house
OxO_12h_v_0h	OXO transgenic 12hpi vs 0hpi	In-house
OxO_24h_v_0h	OXO transgenic 24hpi vs 0hpi	In-house
OxO_36h_v_0h	OXO transgenic 36hpi vs 0hpi	In-house
Chemical treatments		
H2O24vH2O55_AC	OA infiltrated leaves, AC Colibri	In-house
H2O24vH2O55_OxO	OA infiltrated leaves, OXO transgenic	In-house
OA24vH2O24_AC	OA infiltrated leaves, AC Colibri	In-house
OA24vH2O24_OxO	OA infiltrated leaves, OXO transgenic	In-house
BW_T1	Bentazon vs water sprayed leaves, 1 hour post treatment	Zhu et al., 2009
BW_T2	Bentazon vs water sprayed leaves, 2 hour post treatment	Zhu et al., 2009
BW_T4	Bentazon vs water sprayed leaves, 4 hour post treatment	Zhu et al., 2009
BW_T8	Bentazon vs water sprayed leaves, 8 hour post treatment	Zhu et al., 2009
24C	Glyphosate, 24 hour post treatment	Zhu et al., 2008

Table 2.1. (cont.)

4C	Glyphosate, 4 hour post treatment	Zhu et al., 2008
1C	Glyphosate, 1 hour post treatment	Zhu et al., 2008
RtGthT4T0	Soybean root growth 4hpi vs 0hpi	In-house
Hairy roots		
HvN	Hairy root vs normal root	In-house

Table 2.2. 411 soybean genes that show strong response to pathogens.

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma01g01400.1	ACM89637.1	NBS-LRR type disease resistance protein [Glycine max]	1658 bits (4292)	0	847/891 (95%)
Glyma01g02580.1	ACQ90242.1	cinnamoyl alcohol dehydrogenase [Glycine max]	564 bits (1453)	9E-159	278/358 (77%)
Glyma01g03820.1	ACM89738.1	mitochondrial benzaldehyde dehydrogenase[Antirrhinum majus]	874 bits (2258)	7E-252	414/499 (82%)
Glyma01g04360.1	ACU24113.1	unknown [Glycine max]	653 bits (1683)	3E-185	323/344 (93%)
Glyma01g04730.1	NP_174030.2	unknown protein [Arabidopsis thaliana]	470 bits (1208)	4E-130	261/522 (50%)
Glyma01g06820.1	XP_002528475.1	"Leucoanthocyanidin dioxygenase, putative[Ricinus communis]"	409 bits (1050)	5E-112	192/350 (54%)
Glyma01g17590.1	No Hit Found				
Glyma01g24150.2	ACM89535.1	protein kinase [Glycine max]	641 bits (1651)	1E-181	322/412 (78%)
Glyma01g24950.4	ACJ84681.1	unknown [Medicago truncatula]	523 bits (1346)	3E-146	250/313 (79%)
Glyma01g26220.1	ACU16470.1	unknown [Glycine max]	424 bits (1089)	2E-116	207/219 (94%)
Glyma01g31600.1	XP_002526684.1	"Prenylated Rab acceptor protein, putative[Ricinus communis]"	269 bits (686)	6E-70	137/203 (67%)
Glyma01g31750.1	ACU19308.1	unknown [Glycine max]	364 bits (932)	1E-98	179/188 (95%)
Glyma01g33150.1	O49859.1	RecName: Full=Cytochrome P450 82A4; AltName:Full=Cytochrome P450 CP9	1042 bits (2694)	2E-302	521/526 (99%)
Glyma01g37810.1	CAA06027.1	NADPH:isoflavone reductase [Glycine max]	629 bits (1622)	3E-178	317/318 (99%)
Glyma01g38150.1	XP_002310223.1	chromatin remodeling complex subunit [Populustrichocarpa]	1103 bits (2851)	0	555/754 (73%)
Glyma01g39260.1	CAA87077.1	heat shock transcription factor 34 [Glycine max]	508 bits (1307)	1E-141	249/282 (88%)
Glyma01g40130.2	AAG28435.1	plasma membrane Ca2+-ATPase [Glycine max]	1779 bits (4606)	0	909/936 (97%)
Glyma01g40320.1	ACU15686.1	unknown [Glycine max]	195 bits (494)	6E-48	115/158 (72%)
Glyma01g41030.2	ACU17885.1	unknown [Glycine max]	479 bits (1231)	5E-133	239/239 (100%)
Glyma01g42390.1	AAW82960.1	senescence-inducible chloroplast stay-green protein2 [Glycine max]	522 bits (1343)	5E-146	257/271 (94%)
Glyma01g42420.1	ACA49724.1	phospholipase D gamma [Citrus sinensis]	1278 bits (3307)	0	607/852 (71%)
Glyma01g42660.1	ACU19846.1	unknown [Glycine max]	460 bits (1183)	1E-127	210/224 (93%)
Glyma01g42670.1	BAH01715.1	PR-5 protein [Glycine max]	497 bits (1279)	1E-138	221/240 (92%)
Glyma01g43420.1	ACU19938.1	unknown [Glycine max]	463 bits (1189)	4E-128	226/226 (100%)
Glyma01g43880.1	P30081.1	RecName: Full=Chalcone synthase 7; AltName:Full=Naringenin-chalcone synthase 7	784 bits (2023)	9E-225	389/389 (100%)
Glyma01g44120.1	ACU13860.1	unknown [Glycine max]	238 bits (607)	8E-61	113/113 (100%)
Glyma01g44600.1	XP_002281119.1	PREDICTED: hypothetical protein [Vitisvinifera]	664 bits (1712)	9E-189	314/366 (85%)
Glyma01g44930.1	CAB52689.1	hexose transporter [Solanum lycopersicum]	835 bits (2156)	5E-240	421/509 (82%)
Glyma01g44930.1	CAB52689.1	hexose transporter [Solanum lycopersicum]	835 bits (2156)	5E-240	421/509 (82%)
Glyma01g45500.1	XP_002510144.1	conserved hypothetical protein [Ricinuscommunis]	44 bits (101)	0.013	25/61 (40%)
Glyma02g00250.1	XP_002280070.1	PREDICTED: hypothetical protein [Vitisvinifera]	679 bits (1751)	5E-193	363/631 (57%)
Glyma02g00820.1	BAA81732.1	GmMYB29A2 [Glycine max]	545 bits (1403)	6E-153	263/264 (99%)
Glyma02g00870.1	ADE41122.1	AP2 domain class transcription factor [Malus xdomestica]	178 bits (451)	1E-42	98/188 (52%)

Table 2.2 (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma02g01100.1	XP_002273987.1	PREDICTED: hypothetical protein [Vitisvinifera]	1842 bits (4770)	0	948/1246 (76%)
Glyma02g02780.1	ACM89628.1	TIR-NBS-LRR type disease resistance protein[Glycine max]	518 bits (1332)	1E-144	257/257 (100%)
Glyma02g03290.1	ABW96008.1	matrix metalloproteinase [Glycine max]	541 bits (1393)	1E-151	273/305 (89%)
Glyma02g11610.1	BAB83692.1	ABA-glucosyltransferase [Vigna angularis]	759 bits (1958)	4E-217	371/477 (77%)
Glyma02g11800.1	BAH03204.1	peroxisomal adenine nucleotide carrier 1 [Glycinemax]	585 bits (1508)	5E-165	305/318 (95%)
Glyma02g15150.1	XP_002518792.1	"catalytic, putative [Ricinus communis]"	251 bits (639)	4E-64	114/183 (62%)
Glyma02g18380.3	ACU24570.1	unknown [Glycine max]	674 bits (1737)	1E-191	334/339 (98%)
Glyma02g26160.1	ACJ54281.1	lipoxygenase [Camellia sinensis]	1135 bits (2935)	0	542/837 (64%)
Glyma02g32000.1	XP_002284060.1	PREDICTED: hypothetical protein [Vitisvinifera]	1096 bits (2833)	0	525/658 (79%)
Glyma02g34940.1	XP_002323366.1	predicted protein [Populus trichocarpa]	416 bits (1069)	7E-114	223/427 (52%)
Glyma02g35230.1	XP_002524068.1	"syntaxin, putative [Ricinus communis]"	396 bits (1016)	5E-108	206/303 (67%)
Glyma02g39870.1	CAP08303.1	DNA-binding protein [Vitis thunbergii]	625 bits (1610)	1E-176	327/516 (63%)
Glyma02g40040.1	AAD37428.1	peroxidase 3 precursor [Phaseolusvulgaris]	503 bits (1293)	3E-140	245/299 (81%)
Glyma02g40290.2	Q42797.1	RecName: Full=Trans-cinnamate4-monooxygenase; AltName: Full=Cinnamic acid 4-hydroxylase; Short=CA4H;Short=C4H; AltName: Full=Cytochrome P450C4H; AltName: Full=CytochromeP450 73	732 bits (1889)	4E-209	364/392 (92%)
Glyma02g40820.4	ACU18009.1	unknown [Glycine max]	799 bits (2061)	4E-229	395/395 (100%)
Glyma02g41490.1	ACU19261.1	unknown [Glycine max]	789 bits (2037)	2E-226	391/392 (99%)
Glyma02g42140.2	ABH02835.1	MYB transcription factor MYB75 [Glycine max]	559 bits (1438)	7E-157	272/306 (88%)
Glyma02g42250.1	ACF37256.1	nematode resistance HS1pro1 protein [Glycine max]	387 bits (993)	1E-105	189/189 (100%)
Glyma02g42730.1	ACU23245.1	unknown [Glycine max]	620 bits (1597)	2E-175	312/326 (95%)
Glyma02g43580.1	ACU24097.1	unknown [Glycine max]	581 bits (1497)	8E-164	288/307 (93%)
Glyma02g45970.3	ACU24054.1	unknown [Glycine max]	691 bits (1781)	8E-197	335/336 (99%)
Glyma02g47210.2	ACI31551.1	heat shock protein 90-2 [Glycine max]	865 bits (2234)	5E-249	443/487 (90%)
Glyma02g47370.1	XP_002533727.1	"amino acid transporter, putative [Ricinuscommunis]"	572 bits (1472)	1E-160	269/451 (59%)
Glyma02g47590.1	ABZ11028.1	Potyvirus VPg interacting protein [Arachishypogaea]	1038 bits (2684)	3E-301	510/567 (89%)
Glyma02g47750.1	P26690.1	RecName: Full=NAD(P)H-dependent6'-deoxychalcone synthase	609 bits (1570)	2E-172	299/315 (94%)
Glyma03g00380.1	BAB86895.1	syringolide-induced protein B15-3-5 [Glycine max]	481 bits (1238)	5E-134	230/230 (100%)
Glyma03g03190.2	ABA54869.1	putative 5-enolpyruvylshikimate 3-phosphate synthase[Fagus sylvatica]	623 bits (1606)	3E-176	325/445 (73%)
Glyma03g16600.1	AAG34803.1	glutathione S-transferase GST 13 [Glycinemax]	409 bits (1050)	5E-112	198/210 (94%)
Glyma03g26060.2	ACU17256.1	unknown [Glycine max]	158 bits (397)	1E-36	84/134 (62%)
Glyma03g27030.1	ACU24346.1	unknown [Glycine max]	809 bits (2087)	5E-232	393/420 (93%)
Glyma03g28850.1	Q03773.1	"RecName: Full=Glucanendo-1,3-beta-glucosidase; AltName: Full=(1->3)-beta-glucanendohydrolase; Short=(1->3)-beta-glucanase; AltName:Full=Beta-1,3-endoglucanase; Flags: Precursor"	695 bits (1791)	7E-198	347/347 (100%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma03g33340.3	ACU24136.1	unknown [Glycine max]	424 bits (1090)	1E-116	205/207 (99%)
Glyma03g33340.4	ACU24136.1	unknown [Glycine max]	479 bits (1232)	4E-133	232/235 (98%)
Glyma03g34480.1	BAB86932.1	glucosyltransferase-14 [Vigna angularis]	583 bits (1502)	4E-164	286/470 (60%)
Glyma03g37390.1	XP_002523112.1	"Pectinesterase-2 precursor, putative [Ricinuscommunis]"	585 bits (1508)	4E-165	285/362 (78%)
Glyma03g37650.1	XP_002275221.1	PREDICTED: hypothetical protein [Vitisvinifera]	599 bits (1543)	5E-169	288/324 (88%)
Glyma03g37940.1	ABC26919.1	WRKY51 [Glycine max]	508 bits (1307)	9E-142	250/285 (87%)
Glyma03g39440.1	P48490.1	RecName: Full=Serine/threonine-proteinphosphatase PP1	612 bits (1576)	6E-173	296/310 (95%)
Glyma03g39610.1	AAW78864.1	respiratory burst oxidase 2 [Medicago truncatula]	1475 bits (3818)	0	731/874 (83%)
Glyma03g41100.1	ABH02843.1	MYB transcription factor MYB88 [Glycine max]	430 bits (1104)	2E-118	206/209 (98%)
Glyma03g41390.1	ACU17303.1	unknown [Glycine max]	173 bits (438)	1E-41	87/87 (100%)
Glyma04g03740.1	XP_002283780.1	PREDICTED: hypothetical protein [Vitisvinifera]	714 bits (1842)	1E-203	339/465 (72%)
Glyma04g04240.1	ABC47847.1	N-hydroxycinnamoyl/benzoyltransferase 5 [Glycinemax]	838 bits (2164)	6E-241	408/423 (96%)
Glyma04g04250.1	ABC47846.1	N-hydroxycinnamoyl/benzoyltransferase 4 [Glycinemax]	969 bits (2504)	2E-280	469/469 (100%)
Glyma04g07980.1	XP_002332113.1	predicted protein [Populus trichocarpa]	457 bits (1174)	3E-126	268/509 (52%)
Glyma04g08020.1	XP_002532121.1	conserved hypothetical protein [Ricinuscommunis]	125 bits (312)	6E-27	52/66 (78%)
Glyma04g14800.3	Q07185.1	"RecName: Full=Alternative oxidase 1,mitochondrial; Flags: Precursor"	375 bits (962)	2E-101	181/183 (98%)
Glyma04g29490.1	ACU15451.1	unknown [Glycine max]	42 bits (97)	0.117	23/39 (58%)
Glyma04g34600.1	XP_002511496.1	"HIPL1 protein precursor, putative [Ricinuscommunis]"	1008 bits (2604)	6E-292	471/630 (74%)
Glyma04g38560.1	ACU20768.1	unknown [Glycine max]	562 bits (1448)	4E-158	271/293 (92%)
Glyma04g39650.1	ABC26913.1	WRKY21 [Glycine max]	382 bits (981)	2E-104	180/196 (91%)
Glyma04g40700.1	ACU15418.1	unknown [Glycine max]	89 bits (219)	2E-16	44/62 (70%)
Glyma04g40860.1	ACU15271.1	unknown [Glycine max]	82 bits (202)	2E-14	40/69 (57%)
Glyma04g41060.1	XP_002318453.1	predicted protein [Populus trichocarpa]	520 bits (1339)	2E-145	275/454 (60%)
Glyma04g42250.1	XP_002520270.1	"protein transporter, putative [Ricinuscommunis]"	631 bits (1625)	2E-178	324/509 (63%)
Glyma04g42360.1	ACU16164.1	unknown [Glycine max]	277 bits (708)	8E-73	139/151 (92%)
Glyma04g42800.1	ACD39369.1	NAC domain protein [Glycine max]	602 bits (1550)	1E-169	286/300 (95%)
Glyma05g00640.1	ACU23055.1	unknown [Glycine max]	417 bits (1071)	2E-114	215/292 (73%)
Glyma05g02210.1	XP_002511397.1	conserved hypothetical protein [Ricinuscommunis]	251 bits (639)	2E-64	138/298 (46%)
Glyma05g04180.2	XP_002509588.1	"lipid binding protein, putative [Ricinuscommunis]"	102 bits (254)	1E-19	46/83 (55%)
Glyma05g06080.3	ACU21254.1	unknown [Glycine max]	502 bits (1292)	6E-140	244/277 (88%)
Glyma05g09440.2	AAN62760.1	disease resistance protein-like protein MsR1[Medicago sativa]	741 bits (1911)	2E-211	381/652 (58%)
Glyma05g17470.1	AAN62760.1	disease resistance protein-like protein MsR1[Medicago sativa]	700 bits (1804)	5E-199	348/530 (65%)
Glyma05g23890.1	ACU15315.1	unknown [Glycine max]	112 bits (280)	4E-23	58/72 (80%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma05g24610.1	ACU22700.1	unknown [Glycine max]	724 bits (1868)	1E-206	352/384 (91%)
Glyma05g26420.1	BAE71282.1	putative receptor-like GPI-anchored protein 2[Trifolium pratense]	394 bits (1011)	2E-107	185/305 (60%)
Glyma05g27960.2	ACU15797.1	unknown [Glycine max]	467 bits (1200)	1E-129	236/236 (100%)
Glyma05g29390.1	AAG34807.1	glutathione S-transferase GST 17 [Glycinemax]	424 bits (1089)	9E-117	207/229 (90%)
Glyma05g29400.1	ACU14737.1	unknown [Glycine max]	413 bits (1059)	2E-113	203/223 (91%)
Glyma05g32850.1	ACD39384.1	NAC domain protein [Glycine max]	630 bits (1624)	1E-178	296/298 (99%)
Glyma05g33050.1	XP_002329453.1	predicted protein [Populus trichocarpa]	797 bits (2058)	1E-228	403/587 (68%)
Glyma05g33650.2	ACJ84884.1	unknown [Medicago truncatula]	545 bits (1404)	6E-153	287/365 (78%)
Glyma05g35500.1	ACU13449.1	unknown [Glycine max]	210 bits (533)	2E-52	101/101 (100%)
Glyma05g36100.1	ACU22894.1	unknown [Glycine max]	657 bits (1694)	1E-186	311/312 (99%)
Glyma05g37590.1	ACU14098.1	unknown [Glycine max]	387 bits (992)	3E-105	185/216 (85%)
Glyma05g37770.2	BAJ10680.1	bHLH transcription factor [Lotus japonicus]	596 bits (1534)	6E-168	318/474 (67%)
Glyma05g38540.3	XP_002284543.1	PREDICTED: hypothetical protein [Vitisvinifera]	1050 bits (2714)	2E-304	548/810 (67%)
Glyma06g00990.1	Q39827.1	RecName: Full=Arginine decarboxylase;Short=ARGDC; Short=ADC	1201 bits (3105)	0	603/693 (87%)
Glyma06g02640.1	XP_002274931.1	PREDICTED: hypothetical protein [Vitisvinifera]	591 bits (1523)	1E-166	315/536 (58%)
Glyma06g05550.1	XP_002278410.1	PREDICTED: hypothetical protein isoform 1[Vitis vinifera]	495 bits (1272)	1E-137	245/334 (73%)
Glyma06g07560.1	XP_002309595.1	predicted protein [Populus trichocarpa]	113 bits (281)	3E-23	64/146 (43%)
Glyma06g07920.2	XP_002283635.1	PREDICTED: hypothetical protein [Vitisvinifera]	1429 bits (3697)	0	716/1093 (65%)
Glyma06g08830.1	ACU20666.1	unknown [Glycine max]	206 bits (523)	2E-51	99/100 (99%)
Glyma06g09220.3	NP_001152396.1	LOC100286036 [Zea mays]	910 bits (2350)	2E-262	439/567 (77%)
Glyma06g11140.1	XP_002285810.1	PREDICTED: hypothetical protein [Vitisvinifera]	1016 bits (2626)	2E-294	518/659 (78%)
Glyma06g11990.1	XP_002306645.1	predicted protein [Populus trichocarpa]	508 bits (1307)	1E-141	234/368 (63%)
Glyma06g13020.1	ACU13806.1	unknown [Glycine max]	314 bits (803)	1E-83	148/148 (100%)
Glyma06g14100.1	XP_002891016.1	hypothetical protein ARALYDRAFT_890872[Arabidopsis lyrata subsp. lyrata]	49 bits (115)	0.0001	22/41 (53%)
Glyma06g14820.4	AAT94362.1	putative chalcone isomerase 4 [Glycine max]	213 bits (540)	5E-53	113/133 (84%)
Glyma06g15030.1	ACU17865.1	unknown [Glycine max]	635 bits (1637)	4E-180	318/320 (99%)
Glyma06g15410.1	AAL69983.1	24 kDa protein SC24 [Glycine max]	421 bits (1082)	5E-116	201/219 (91%)
Glyma06g17060.3	Q39817.1	RecName: Full=Calnexin homolog; Flags:Precursor	810 bits (2092)	1E-232	396/438 (90%)
Glyma06g19810.1	ACU15344.1	unknown [Glycine max]	122 bits (304)	3E-26	57/69 (82%)
Glyma06g19890.1	AAX73302.1	EDS1 [Solanum lycopersicum]	567 bits (1460)	3E-159	303/608 (49%)
Glyma06g26610.1	XP_002308977.1	predicted protein [Populus trichocarpa]	85 bits (209)	7E-15	55/160 (34%)
Glyma06g40330.1	ACU16073.1	unknown [Glycine max]	171 bits (432)	9E-41	85/106 (80%)
Glyma06g40710.1	ACM89623.1	TIR-NBS-LRR type disease resistance protein[Glycine max]	1689 bits (4373)	0	876/1112 (78%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma06g40740.2	ACM89623.1	TIR-NBS-LRR type disease resistance protein[Glycine max]	868 bits (2241)	1E-249	502/822 (61%)
Glyma06g40790.1	ABN08816.1	Aldo/keto reductase [Medicago truncatula]	554 bits (1426)	1E-155	271/336 (80%)
Glyma06g42170.1	XP_002519016.1	"Polyphenol oxidase, chloroplast precursor,putative [Ricinus communis]"	649 bits (1672)	6E-184	329/588 (55%)
Glyma06g44870.2	CBI21830.1	unnamed protein product [Vitis vinifera]	136 bits (341)	1E-29	81/220 (36%)
Glyma06g46190.1	XP_002524184.1	"aconitase, putative [Ricinus communis]"	1628 bits (4214)	0	795/881 (90%)
Glyma06g46680.1	BAA85654.1	hsr203J homolog [Pisum sativum]	536 bits (1379)	4E-150	254/339 (74%)
Glyma06g47560.1	XP_002534670.1	"Derlin-2, putative [Ricinus communis]"	421 bits (1080)	1E-115	199/277 (71%)
Glyma07g02180.2	XP_002279139.1	PREDICTED: hypothetical protein [Vitisvinifera]	757 bits (1952)	3E-216	371/481 (77%)
Glyma07g04470.1	ABC59081.1	cytochrome P450 monooxygenase CYP92A29 [Medicagotruncatula]	828 bits (2137)	8E-238	395/479 (82%)
Glyma07g04480.1	CAC43237.1	lipoxygenase [Sesbania rostrata]	1561 bits (4041)	0	770/925 (83%)
Glyma07g05480.1	ABD32716.1	"O-methyltransferase, family 2; Dimerisation[Medicago truncatula]"	466 bits (1199)	3E-129	232/377 (61%)
Glyma07g06120.2	XP_002523344.1	protein with unknown function [Ricinuscommunis]	749 bits (1933)	4E-214	381/565 (67%)
Glyma07g06150.1	CBI40530.3	unnamed protein product [Vitis vinifera]	433 bits (1113)	4E-119	206/353 (58%)
Glyma07g08310.1	ACU19695.1	unknown [Glycine max]	146 bits (366)	6E-33	78/132 (59%)
Glyma07g10690.1	ACM89469.1	serine/threonine protein kinase family protein[Glycine max]	599 bits (1542)	1E-168	339/627 (54%)
Glyma07g16810.1	P32110.1	RecName: Full=Probable glutathioneS-transferase; AltName: Full=Heat shock protein 26A; AltName:Full=G2-4	455 bits (1170)	4E-126	225/225 (100%)
Glyma07g16830.1	ACU15816.1	unknown [Glycine max]	449 bits (1154)	3E-124	221/224 (98%)
Glyma07g16910.1	AAG34798.1	glutathione S-transferase GST 8 [Glycinemax]	455 bits (1170)	5E-126	224/225 (99%)
Glyma07g30290.1	P37900.1	"RecName: Full=Heat shock 70 kDa protein,mitochondrial; Flags: Precursor"	1157 bits (2991)	0	594/647 (91%)
Glyma07g30880.1	AAB06594.1	sugar transporter [Medicago truncatula]	859 bits (2218)	3E-247	425/517 (82%)
Glyma07g32340.1	AAF45142.1	isoflavone synthase 1 [Glycine max]	456 bits (1171)	9E-126	239/278 (85%)
Glyma07g33870.1	ACU24016.1	unknown [Glycine max]	399 bits (1024)	5E-109	206/222 (92%)
Glyma07g35370.1	XP_002316103.1	predicted protein [Populus trichocarpa]	815 bits (2104)	1E-233	494/1091 (45%)
Glyma07g36160.1	CAN65735.1	hypothetical protein [Vitis vinifera]	1318 bits (3410)	0	664/1117 (59%)
Glyma07g37250.1	P26987.1	RecName: Full=Stress-induced protein SAM22;AltName: Full=Starvation-associated message 22; AltName: Allergen=Gly m4	316 bits (808)	3E-84	158/158 (100%)
Glyma07g37270.1	ACU16890.1	unknown [Glycine max]	318 bits (813)	7E-85	158/158 (100%)
Glyma07g39130.1	AAL66290.1	adenosine 5'-phosphosulfate reductase[Glycine max]	743 bits (1918)	2E-212	378/469 (80%)
Glyma07g39630.1	ABN08458.1	Cytochrome b561 / ferric reductase transmembrane[Medicago truncatula]	455 bits (1169)	1E-125	217/302 (71%)
Glyma08g02580.1	ABS18424.1	WRKY20 [Glycine max]	530 bits (1365)	3E-148	255/268 (95%)
Glyma08g02840.2	ACU17761.1	unknown [Glycine max]	374 bits (959)	3E-101	189/269 (70%)
Glyma08g06340.1	ABD32592.1	At1g61340 [Medicago truncatula]	138 bits (345)	2E-30	77/156 (49%)
Glyma08g07690.1	AAP68983.1	alternative oxidase 2b [Glycine max]	628 bits (1618)	7E-178	306/326 (93%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma08g08380.1	CAI99394.1	polygalacturonase inhibiting protein precursor[Glycine max]	633 bits (1631)	2E-179	313/313 (100%)
Glyma08g08860.1	ACU16204.1	unknown [Glycine max]	107 bits (265)	1E-21	50/50 (100%)
Glyma08g11610.1	ABQ63059.1	chalcone synthase 9 [Glycine max]	756 bits (1950)	3E-216	375/375 (100%)
Glyma08g11620.1	P24826.1	RecName: Full=Chalcone synthase 1; AltName:Full=Naringenin-chalcone synthase 1	778 bits (2009)	4E-223	388/388 (100%)
Glyma08g11980.1	XP_002285701.1	PREDICTED: hypothetical protein [Vitisvinifera]	1074 bits (2775)	0	534/719 (74%)
Glyma08g12530.1	ACU14737.1	unknown [Glycine max]	368 bits (944)	1E-99	179/221 (80%)
Glyma08g12630.1	AAL12248.1	heat shock transcription factor [Phaseolusacutifolius]	693 bits (1786)	3E-197	334/402 (83%)
Glyma08g13510.1	ACU13882.1	unknown [Glycine max]	251 bits (640)	7E-65	121/121 (100%)
Glyma08g16050.1	XP_002513196.1	conserved hypothetical protein [Ricinuscommunis]	49 bits (114)	0.0002	28/63 (44%)
Glyma08g16770.2	AAR88248.1	mitochondrial citrate synthase precursor [Citrusjunos]	691 bits (1781)	1E-196	332/462 (71%)
Glyma08g18690.2	AAG34800.1	glutathione S-transferase GST 10 [Glycinemax]	219 bits (557)	5E-55	105/105 (100%)
Glyma08g18990.1	XP_002272947.1	PREDICTED: hypothetical protein [Vitisvinifera]	513 bits (1319)	3E-143	247/329 (75%)
Glyma08g19250.1	CBI19107.1	unnamed protein product [Vitis vinifera]	636 bits (1638)	5E-180	304/438 (69%)
Glyma08g21840.2	XP_002279139.1	PREDICTED: hypothetical protein [Vitisvinifera]	584 bits (1505)	2E-164	289/376 (76%)
Glyma08g23380.4	ABY84656.1	transcription factor [Glycine max]	496 bits (1276)	4E-138	255/313 (81%)
Glyma08g25480.1	XP_002264007.1	PREDICTED: hypothetical protein [Vitisvinifera]	370 bits (948)	3E-100	197/280 (70%)
Glyma08g26110.1	ACU16880.1	unknown [Glycine max]	275 bits (703)	4E-72	139/157 (88%)
Glyma08g26830.1	ACU24307.1	unknown [Glycine max]	514 bits (1323)	1E-143	269/466 (57%)
Glyma08g27070.1	ACU18983.1	unknown [Glycine max]	647 bits (1667)	1E-183	315/333 (94%)
Glyma08g37670.1	ABF18929.1	putative 1-deoxy-D-xylulose 5-phosphate synthase[Hevea brasiliensis]	897 bits (2317)	1E-258	445/505 (88%)
Glyma08g42050.1	XP_002520371.1	"basic 7S globulin 2 precursor small subunit,putative [Ricinus communis]"	505 bits (1300)	8E-141	237/317 (74%)
Glyma08g42070.2	ACU21363.1	unknown [Glycine max]	734 bits (1894)	8E-210	368/389 (94%)
Glyma08g45520.1	ACU16888.1	unknown [Glycine max]	420 bits (1078)	1E-115	207/207 (100%)
Glyma08g46650.1	XP_002316674.1	predicted protein [Populus trichocarpa]	314 bits (804)	4E-83	150/238 (63%)
Glyma08g47620.1	XP_002302888.1	predicted protein [Populus trichocarpa]	496 bits (1276)	3E-138	237/291 (81%)
Glyma09g00280.1	XP_002265125.1	PREDICTED: hypothetical protein [Vitisvinifera]	703 bits (1813)	2E-200	332/391 (84%)
Glyma09g00820.1	ABS18425.1	WRKY23 [Glycine max]	846 bits (2184)	3E-243	425/493 (86%)
Glyma09g02650.1	ACU21261.1	unknown [Glycine max]	601 bits (1547)	1E-169	306/347 (88%)
Glyma09g03490.3	XP_002529901.1	"UDP-glucose 4-epimerase, putative [Ricinuscommunis]"	725 bits (1871)	4E-207	354/410 (86%)
Glyma09g04520.1	ACU16114.1	unknown [Glycine max]	310 bits (793)	2E-82	156/157 (99%)
Glyma09g05700.1	ACU20253.1	unknown [Glycine max]	87 bits (215)	2E-15	46/104 (44%)
Glyma09g07330.1	XP_002300900.1	predicted protein [Populus trichocarpa]	581 bits (1496)	1E-163	268/417 (64%)
Glyma09g23140.1	BAA77675.1	Chitinase III-A [Glycine max]	304 bits (777)	2E-80	148/182 (81%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma09g24170.1	ACU14624.1	unknown [Glycine max]	165 bits (416)	1E-38	84/110 (76%)
Glyma09g24410.1	XP_002513649.1	"heat shock protein, putative [Ricinuscommunis]"	1137 bits (2940)	0	588/699 (84%)
Glyma09g25470.4	XP_002267459.1	PREDICTED: hypothetical protein [Vitisvinifera]	619 bits (1595)	5E-175	310/417 (74%)
Glyma09g32890.1	XP_002268615.1	PREDICTED: hypothetical protein [Vitisvinifera]	545 bits (1404)	5E-153	278/472 (58%)
Glyma09g34090.1	ACU17679.1	unknown [Glycine max]	501 bits (1289)	1E-139	254/295 (86%)
Glyma09g38410.2	ACU19663.1	unknown [Glycine max]	716 bits (1848)	2E-204	336/358 (93%)
Glyma09g38550.1	AAQ84167.1	isopentenyl pyrophosphate isomerase [Puerariamontana var. lobata]	495 bits (1272)	1E-137	249/302 (82%)
Glyma09g41450.1	P22195.2	RecName: Full=Cationic peroxidase 1; AltName:Full=PNPC1; Flags: Precursor	491 bits (1263)	9E-137	244/316 (77%)
Glyma09g41850.1	ACU18726.1	unknown [Glycine max]	722 bits (1863)	3E-206	356/357 (99%)
Glyma10g00470.1	ACU16024.1	unknown [Glycine max]	297 bits (760)	1E-78	150/150 (100%)
Glyma10g01640.1	ABI34647.1	bZIP transcription factor bZIP59 [Glycine max]	300 bits (766)	4E-80	152/152 (100%)
Glyma10g02210.1	ACU14506.1	unknown [Glycine max]	181 bits (459)	7E-44	92/92 (100%)
Glyma10g02370.2	XP_002523063.1	"multidrug resistance-associated protein 2, 6(mrp2, 6), abc-transoprter, putative [Ricinus communis]"	1943 bits (5031)	0	965/1382 (69%)
Glyma10g05210.1	ACU15884.1	unknown [Glycine max]	258 bits (657)	7E-67	143/185 (77%)
Glyma10g05830.1	CAA45621.1	polyubiquitin [Petroselinum crispum]	593 bits (1528)	2E-167	306/306 (100%)
Glyma10g06600.1	ACS88364.1	phenylalanine ammonia-lyase 2 [Glycine max]	1415 bits (3662)	0	716/717 (99%)
Glyma10g07050.1	ACN97419.1	thaumatin-like protein [Pyrus pyrifolia]	387 bits (992)	2E-105	177/235 (75%)
Glyma10g28830.1	ACU20148.1	unknown [Glycine max]	421 bits (1082)	1E-115	218/267 (81%)
Glyma10g29210.1	ACU20489.1	unknown [Glycine max]	149 bits (376)	3E-34	73/90 (81%)
Glyma10g29280.1	AAW78864.1	respiratory burst oxidase 2 [Medicago truncatula]	1437 bits (3718)	0	698/829 (84%)
Glyma10g31590.1	XP_002518910.1	"cystathionine gamma-synthase, putative[Ricinus communis]"	653 bits (1683)	3E-185	328/447 (73%)
Glyma10g32190.1	AAA34015.1	calmodulin [Glycine max]	296 bits (756)	4E-78	150/150 (100%)
Glyma10g32820.1	ACU19481.1	unknown [Glycine max]	148 bits (373)	4E-34	70/70 (100%)
Glyma10g33650.1	AAG34805.1	glutathione S-transferase GST 15 [Glycinemax]	460 bits (1182)	1E-127	221/221 (100%)
Glyma10g36680.1	BAA01950.1	peroxidase [Vigna angularis]	599 bits (1542)	3E-169	294/344 (85%)
Glyma10g41260.1	ACU20604.1	unknown [Glycine max]	288 bits (736)	7E-76	149/177 (84%)
Glyma10g42170.1	ACU24031.1	unknown [Glycine max]	258 bits (658)	9E-67	128/159 (80%)
Glyma10g42520.1	ACU13850.1	unknown [Glycine max]	151 bits (379)	1E-34	78/111 (70%)
Glyma10g42820.3	XP_002521290.1	conserved hypothetical protein [Ricinuscommunis]	196 bits (498)	7E-48	119/275 (43%)
Glyma10g43760.3	XP_002879795.1	esterase/lipase/thioesterase family protein[Arabidopsis lyrata subsp. lyrata]	449 bits (1153)	6E-124	202/292 (69%)
Glyma10g44160.1	ACU24004.1	unknown [Glycine max]	574 bits (1479)	8E-162	291/325 (89%)
Glyma11g00230.1	ACU22845.1	unknown [Glycine max]	616 bits (1586)	5E-174	301/303 (99%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma11g00710.1	CAB52689.1	hexose transporter [Solanum lycopersicum]	846 bits (2185)	2E-243	430/524 (82%)
Glyma11g01240.1	AAC97389.1	4-coumarate:CoA ligase isoenzyme 3 [Glycine max]	667 bits (1719)	2E-189	342/357 (95%)
Glyma11g01350.2	AAO67373.1	chalcone synthase [Glycine max]	664 bits (1713)	8E-189	330/331 (99%)
Glyma11g02770.1	ACU19846.1	unknown [Glycine max]	484 bits (1245)	9E-135	224/224 (100%)
Glyma11g02980.1	AAW82959.1	senescence-inducible chloroplast stay-green protein1 [Glycine max]	508 bits (1306)	1E-141	247/261 (94%)
Glyma11g04040.2	ACU19401.1	unknown [Glycine max]	429 bits (1101)	6E-118	219/237 (92%)
Glyma11g04390.8	ACU19472.1	unknown [Glycine max]	465 bits (1196)	4E-129	233/234 (99%)
Glyma11g05300.2	ACU23656.1	unknown [Glycine max]	359 bits (919)	8E-97	176/176 (100%)
Glyma11g05530.1	CAB43505.1	cytochrome P450 [Cicer arietinum]	674 bits (1739)	9E-192	331/463 (71%)
Glyma11g08920.1	XP_002528761.1	"isocitrate dehydrogenase, putative [Ricinuscommunis]"	614 bits (1581)	1E-173	306/350 (87%)
Glyma11g10230.1	ACU18016.1	unknown [Glycine max]	562 bits (1446)	9E-158	274/302 (90%)
Glyma11g13270.1	ACU16135.1	unknown [Glycine max]	464 bits (1193)	8E-129	220/235 (93%)
Glyma11g15610.1	XP_002725583.1	PREDICTED: rCG54380-like [Rattus norvegicus]	37 bits (84)	0.79	16/34 (47%)
Glyma11g18320.1	ACU24451.1	unknown [Glycine max]	646 bits (1664)	4E-183	331/360 (91%)
Glyma11g20130.1	XP_002534208.1	conserved hypothetical protein [Ricinuscommunis]	1017 bits (2628)	1E-294	530/712 (74%)
Glyma11g21640.1	No Hit Found				
Glyma11g29720.1	CBI15865.3	unnamed protein product [Vitis vinifera]	525 bits (1352)	9E-147	282/462 (61%)
Glyma11g31840.1	XP_002530555.1	conserved hypothetical protein [Ricinuscommunis]	218 bits (553)	1E-54	123/253 (48%)
Glyma11g33040.1	ACU13872.1	unknown [Glycine max]	361 bits (924)	1E-97	185/235 (78%)
Glyma11g33090.2	XP_002305924.1	predicted protein [Populus trichocarpa]	618 bits (1593)	1E-174	305/373 (81%)
Glyma11g33560.3	AAG24873.1	cytosolic glutamine synthetase GSbeta1[Glycine max]	508 bits (1306)	1E-141	238/238 (100%)
Glyma11g36200.1	ACU20263.1	unknown [Glycine max]	810 bits (2090)	2E-232	413/433 (95%)
Glyma11g37930.1	ACU19213.1	unknown [Glycine max]	208 bits (529)	8E-52	104/141 (73%)
Glyma12g01420.1	XP_002274076.1	PREDICTED: hypothetical protein [Vitisvinifera]	405 bits (1040)	2E-110	254/674 (37%)
Glyma12g02520.1	CAN81488.1	hypothetical protein [Vitis vinifera]	795 bits (2051)	2E-227	498/1205 (41%)
Glyma12g02540.1	AAX85983.1	NAC6 protein [Glycine max]	559 bits (1439)	3E-157	269/294 (91%)
Glyma12g03680.1	XP_002277797.1	PREDICTED: hypothetical protein [Vitisvinifera]	576 bits (1483)	5E-162	309/565 (54%)
Glyma12g05080.1	ACU16055.1	unknown [Glycine max]	241 bits (613)	2E-61	116/144 (80%)
Glyma12g05920.1	ACU19646.1	unknown [Glycine max]	843 bits (2176)	2E-242	431/487 (88%)
Glyma12g07100.1	XP_002864651.1	hypothetical protein ARALYDRAFT_496114[Arabidopsis lyrata subsp. lyrata]	387 bits (993)	4E-105	218/413 (52%)
Glyma12g07330.1	ACU20998.1	unknown [Glycine max]	631 bits (1626)	9E-179	297/361 (82%)
Glyma12g08980.1	CAN69750.1	hypothetical protein [Vitis vinifera]	507 bits (1305)	2E-141	245/356 (68%)
Glyma12g13290.1	XP_002517876.1	"protein phosphatase 2c, putative [Ricinuscommunis]"	436 bits (1120)	3E-120	212/283 (74%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma12g15620.1	ABW76287.1	beta-glucosidase G2 [Medicago truncatula]	804 bits (2075)	1E-230	371/491 (75%)
Glyma12g28990.1	NP_001058242.1	Os06g0653900 [Oryza sativa Japonica Group]	107 bits (265)	2E-21	55/69 (79%)
Glyma12g31070.1	ACU14400.1	unknown [Glycine max]	429 bits (1101)	4E-118	206/213 (96%)
Glyma12g33440.1	ACU17275.1	unknown [Glycine max]	50 bits (119)	0.00004	22/23 (95%)
Glyma12g33510.1	ACU23720.1	unknown [Glycine max]	294 bits (750)	2E-77	144/160 (90%)
Glyma12g35380.1	ACU16833.1	unknown [Glycine max]	318 bits (814)	2E-84	171/245 (69%)
Glyma1332s00200.1	ACU20492.1	unknown [Glycine max]	349 bits (895)	2E-94	170/191 (89%)
Glyma13g00380.1	ABC26916.1	WRKY13 [Glycine max]	463 bits (1191)	2E-128	238/324 (73%)
Glyma13g01420.1	XP_002533178.1	"trehalose-6-phosphate synthase, putative[Ricinus communis]"	1096 bits (2833)	0	523/720 (72%)
Glyma13g03990.1	CBI24343.3	unnamed protein product [Vitis vinifera]	509 bits (1309)	7E-142	269/399 (67%)
Glyma13g04670.1	O49858.1	RecName: Full=Cytochrome P450 82A3; AltName:Full=Cytochrome P450 CP6	1007 bits (2601)	1E-291	500/527 (94%)
Glyma13g06230.1	BAF73620.1	malonyl-CoA:isoflavone7-O-glucoside-6''-O-malonyltransferase [Glycine max]	951 bits (2456)	9E-275	467/467 (100%)
Glyma13g16940.1	XP_002516287.1	"12-oxophytodienoate reductase opr, putative[Ricinus communis]"	658 bits (1697)	6E-187	315/387 (81%)
Glyma13g16950.1	XP_002516287.1	"12-oxophytodienoate reductase opr, putative[Ricinus communis]"	596 bits (1534)	4E-168	289/376 (76%)
Glyma13g19730.1	XP_002274246.1	PREDICTED: hypothetical protein [Vitisvinifera]	678 bits (1749)	1E-192	363/545 (66%)
Glyma13g20200.2	CAA45621.1	polyubiquitin [Petroselinum crispum]	302 bits (771)	5E-80	156/163 (95%)
Glyma13g22470.1	XP_002523538.1	conserved hypothetical protein [Ricinuscommunis]	63 bits (151)	0.00000007	38/80 (47%)
Glyma13g22540.1	ACU24578.1	unknown [Glycine max]	235 bits (597)	7E-60	116/141 (82%)
Glyma13g22650.1	XP_002285304.1	PREDICTED: hypothetical protein [Vitisvinifera]	144 bits (363)	3E-32	66/110 (60%)
Glyma13g25570.1	ACU19960.1	unknown [Glycine max]	734 bits (1894)	7E-210	358/367 (97%)
Glyma13g25620.1	BAE71198.1	putative transporter-like protein [Trifoliumpratense]	730 bits (1884)	2E-208	364/492 (73%)
Glyma13g26630.2	ACU13165.1	unknown [Glycine max]	250 bits (638)	3E-64	127/142 (89%)
Glyma13g27820.2	ACJ83356.1	unknown [Medicago truncatula]	227 bits (577)	4E-57	106/144 (73%)
Glyma13g29350.5	XP_002532981.1	conserved hypothetical protein [Ricinuscommunis]	226 bits (575)	5E-57	112/193 (58%)
Glyma13g29580.1	P26413.1	RecName: Full=Heat shock 70 kDa protein	466 bits (1197)	7E-129	233/367 (63%)
Glyma13g29760.1	AAL12248.1	heat shock transcription factor [Phaseolusacutifolius]	557 bits (1434)	2E-156	270/397 (68%)
Glyma13g30990.1	ACE76905.1	ethylene-responsive element binding factor 4[Glycine max]	330 bits (844)	3E-88	161/222 (72%)
Glyma13g31390.1	Q41112.1	RecName: Full=Stress-related protein;AltName: Full=PvSRP	314 bits (804)	1E-83	156/167 (93%)
Glyma13g31920.2	ACU19469.1	unknown [Glycine max]	593 bits (1528)	2E-167	297/314 (94%)
Glyma13g32320.1	XP_002518591.1	"transcription factor, putative [Ricinuscommunis]"	266 bits (679)	8E-69	152/222 (68%)
Glyma13g34100.1	AAU10526.1	putative receptor-like protein kinase 2 [Glycinemax]	1820 bits (4714)	0	919/999 (91%)
Glyma13g35550.1	ACD39382.1	NAC domain protein [Glycine max]	703 bits (1813)	2E-200	335/343 (97%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma13g36110.1	O81972.1	RecName: Full=Cytochrome P450 82A2; AltName:Full=Cytochrome P450 CP4	1007 bits (2601)	1E-291	504/522 (96%)
Glyma13g37000.1	ACU17275.1	unknown [Glycine max]	51 bits (120)	0.0001	23/23 (100%)
Glyma13g38790.4	ACU23708.1	unknown [Glycine max]	286 bits (731)	7E-75	143/163 (87%)
Glyma13g43870.1	XP_002280686.1	PREDICTED: hypothetical protein [Vitisvinifera]	2156 bits (5585)	0	1056/1396 (75%)
Glyma13g43870.3	XP_002280686.1	PREDICTED: hypothetical protein [Vitisvinifera]	2068 bits (5357)	0	1021/1343 (76%)
Glyma13g43870.5	CAN61052.1	hypothetical protein [Vitis vinifera]	1639 bits (4243)	0	827/1153 (71%)
Glyma13g44700.1	ACU19302.1	unknown [Glycine max]	617 bits (1589)	2E-174	305/331 (92%)
Glyma13g44730.1	ABC26917.1	WRKY27 [Glycine max]	441 bits (1134)	1E-121	230/309 (74%)
Glyma1454s00200.2	ACU19732.1	unknown [Glycine max]	504 bits (1296)	8E-141	254/264 (96%)
Glyma14g00870.1	P26690.1	RecName: Full=NAD(P)H-dependent6'-deoxychalcone synthase	424 bits (1090)	1E-116	209/209 (100%)
Glyma14g01960.1	XP_002274103.1	PREDICTED: hypothetical protein [Vitisvinifera]	664 bits (1711)	2E-188	344/548 (62%)
Glyma14g01990.2	CAN68018.1	hypothetical protein [Vitis vinifera]	804 bits (2074)	2E-230	412/667 (61%)
Glyma14g02060.1	ACU23860.1	unknown [Glycine max]	616 bits (1586)	3E-174	300/301 (99%)
Glyma14g02530.3	XP_002529920.1	"dihydrolipoamide succinyltransferase componentof 2-oxoglutarate dehydrogenase, putative [Ricinus communis]"	564 bits (1451)	2E-158	312/471 (66%)
Glyma14g05350.1	ACU20224.1	unknown [Glycine max]	574 bits (1477)	1E-161	285/305 (93%)
Glyma14g08000.1	ACU23997.1	unknown [Glycine max]	508 bits (1307)	5E-142	245/257 (95%)
Glyma14g08910.1	CBI19558.1	unnamed protein product [Vitis vinifera]	589 bits (1516)	6E-166	285/419 (68%)
Glyma14g11150.1	XP_002516151.1	conserved hypothetical protein [Ricinuscommunis]	300 bits (766)	5E-79	141/189 (74%)
Glyma14g35710.1	ACU14922.1	unknown [Glycine max]	263 bits (671)	3E-68	132/132 (100%)
Glyma14g38010.1	XP_002272040.1	PREDICTED: similar to DNA-binding protein[Vitis vinifera]	668 bits (1721)	1E-189	354/582 (60%)
Glyma14g39160.3	S33612	isocitrate dehydrogenase (NADP) (EC 1.1.1.42) - soybean	800 bits (2066)	1E-229	396/396 (100%)
Glyma14g40170.1	ACU20956.1	unknown [Glycine max]	682 bits (1758)	5E-194	336/353 (95%)
Glyma14g40320.4	XP_002510550.1	"endoplasmin, putative [Ricinus communis]"	1196 bits (3092)	0	613/755 (81%)
Glyma15g01690.2	XP_002304183.1	predicted protein [Populus trichocarpa]	433 bits (1112)	3E-119	192/302 (63%)
Glyma15g02400.2	XP_002515946.1	"phosphofructokinase, putative [Ricinuscommunis]"	755 bits (1948)	7E-216	371/500 (74%)
Glyma15g06010.1	XP_002273316.1	PREDICTED: hypothetical protein [Vitisvinifera]	399 bits (1024)	5E-109	185/271 (68%)
Glyma15g06120.1	XP_002267374.1	PREDICTED: hypothetical protein [Vitisvinifera]	773 bits (1994)	3E-221	392/486 (80%)
Glyma15g06530.1	Q01899.1	"RecName: Full=Heat shock 70 kDa protein,mitochondrial; Flags: Precursor"	1180 bits (3051)	0	612/646 (94%)
Glyma15g06780.1	ACU15426.1	unknown [Glycine max]	283 bits (723)	2E-74	129/164 (78%)
Glyma15g06790.1	ACU15426.1	unknown [Glycine max]	294 bits (752)	9E-78	140/164 (85%)
Glyma15g08520.1	ABN08731.1	Thioredoxin fold [Medicago truncatula]	166 bits (418)	5E-39	85/121 (70%)
Glyma15g12240.1	XP_002531605.1	"dopamine beta-monooxygenase, putative [Ricinuscommunis]"	431 bits (1107)	3E-118	207/383 (54%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma15g13500.1	AAC98519.1	peroxidase precursor [Glycine max]	655 bits (1689)	4E-186	330/354 (93%)
Glyma15g19840.4	ACU18981.1	unknown [Glycine max]	359 bits (921)	4E-97	182/193 (94%)
Glyma15g27660.1	ACU14216.1	unknown [Glycine max]	318 bits (813)	1E-84	155/206 (75%)
Glyma15g35410.1	ACU19928.1	unknown [Glycine max]	728 bits (1878)	5E-208	352/361 (97%)
Glyma15g40200.1	AAC18566.1	"2,4-D inducible glutathione S-transferase [Glycinemax]"	404 bits (1037)	1E-110	196/219 (89%)
Glyma15g40290.1	P46417.1	RecName: Full=Glutathione S-transferase 3	430 bits (1105)	2E-118	208/219 (94%)
Glyma15g40760.1	ABD32591.1	conserved hypothetical protein [Medicagotruncatula]	116 bits (290)	1E-24	52/77 (67%)
Glyma15g41840.1	ACU23850.1	unknown [Glycine max]	755 bits (1947)	5E-216	368/369 (99%)
Glyma15g42490.1	XP_002298157.1	predicted protein [Populus trichocarpa]	57 bits (137)	0.000001	31/62 (50%)
Glyma15g42780.1	No Hit Found				
Glyma16g01060.1	ABC59081.1	cytochrome P450 monooxygenase CYP92A29 [Medicagotruncatula]	847 bits (2186)	2E-243	402/480 (83%)
Glyma16g01220.1	ACU24513.1	unknown [Glycine max]	426 bits (1094)	4E-117	219/230 (95%)
Glyma16g01940.3	XP_002872886.1	hypothetical protein ARALYDRAFT_327634[Arabidopsis lyrata subsp. lyrata]	96 bits (237)	1E-17	51/114 (44%)
Glyma16g23870.2	AAX14494.1	calcium-dependent protein kinase CDPK1444 [Medicagotruncatula]	898 bits (2319)	8E-259	452/535 (84%)
Glyma16g25080.1	ACM89629.1	resistance protein [Glycine max]	725 bits (1869)	2E-206	361/388 (93%)
Glyma16g25110.1	ACM89629.1	resistance protein [Glycine max]	574 bits (1477)	3E-161	311/433 (71%)
Glyma16g27880.1	ACU18907.1	unknown [Glycine max]	666 bits (1717)	2E-189	333/345 (96%)
Glyma16g27950.1	BAG50064.1	transcription factor AP2-EREBP [Lotus japonicus]	156 bits (393)	1E-35	83/137 (60%)
Glyma16g29650.1	ACU16811.1	unknown [Glycine max]	207 bits (525)	2E-51	106/130 (81%)
Glyma16g32330.1	AAQ02703.1	CBF-like protein [Glycine max]	328 bits (839)	1E-87	163/236 (69%)
Glyma16g34060.2	AAO23068.1	R 6 protein [Glycine max]	443 bits (1139)	2E-122	220/221 (99%)
Glyma17g01870.1	XP_002265911.1	PREDICTED: hypothetical protein [Vitisvinifera]	731 bits (1887)	7E-209	349/508 (68%)
Glyma17g03340.1	ACU14497.1	unknown [Glycine max]	314 bits (803)	1E-83	156/157 (99%)
Glyma17g03350.2	ACU20716.1	unknown [Glycine max]	282 bits (719)	6E-74	139/157 (88%)
Glyma17g03350.4	ACU16140.1	unknown [Glycine max]	313 bits (801)	2E-83	157/157 (100%)
Glyma17g03910.1	ACU16284.1	unknown [Glycine max]	296 bits (757)	3E-78	145/156 (92%)
Glyma17g04340.1	ABY25855.1	S-adenosylmethionine synthetase [Glycine soja]	770 bits (1986)	2E-220	378/392 (96%)
Glyma17g04710.1	CBI37053.3	unnamed protein product [Vitis vinifera]	167 bits (421)	5E-39	105/289 (36%)
Glyma17g04810.1	ACU20770.1	unknown [Glycine max]	550 bits (1415)	2E-154	279/321 (86%)
Glyma17g06120.1	BAE93461.1	diacylglycerolacyltransferase-1b [Glycine max]	999 bits (2582)	2E-289	486/504 (96%)
Glyma17g06150.2	XP_002279373.1	PREDICTED: hypothetical protein [Vitisvinifera]	502 bits (1290)	1E-139	259/387 (66%)
Glyma17g07190.2	AAL98709.1	4-coumarate:coenzyme A ligase [Glycinemax]	1070 bits (2766)	0	543/546 (99%)
Glyma17g09500.1	ACU14069.1	unknown [Glycine max]	336 bits (860)	4E-90	164/166 (98%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma17g12150.1	AAF66242.1	dicyanin [Solanum lycopersicum]	130 bits (325)	3E-28	66/133 (49%)
Glyma17g14620.1	ACM78616.1	protease inhibitor/seed storage/lipid transferprotein family protein [Tamarix hispida]	113 bits (282)	3E-23	59/139 (42%)
Glyma17g16440.1	ACU15315.1	unknown [Glycine max]	128 bits (320)	6E-28	66/72 (91%)
Glyma17g17810.1	ACU24410.1	unknown [Glycine max]	567 bits (1461)	1E-159	279/293 (95%)
Glyma17g18570.1	ACU15291.1	unknown [Glycine max]	348 bits (891)	8E-94	160/160 (100%)
Glyma17g23740.1	ACU21415.1	unknown [Glycine max]	453 bits (1165)	2E-125	217/217 (100%)
Glyma17g29190.1	ABS18446.1	WRKY50 [Glycine max]	372 bits (953)	9E-101	204/323 (63%)
Glyma17g32750.1	XP_002267722.1	PREDICTED: hypothetical protein [Vitisvinifera]	493 bits (1267)	5E-137	251/431 (58%)
Glyma17g38190.1	XP_002510725.1	"hydroxysteroid dehydrogenase, putative[Ricinus communis]"	676 bits (1743)	4E-192	326/468 (69%)
Glyma18g04660.1	ACU23125.1	unknown [Glycine max]	714 bits (1842)	9E-204	340/342 (99%)
Glyma18g04770.1	BAB86896.1	syringolide-induced protein 13-1-1 [Glycine max]	815 bits (2103)	5E-234	414/431 (96%)
Glyma18g05160.1	ACU14211.1	unknown [Glycine max]	277 bits (707)	7E-73	139/194 (71%)
Glyma18g06200.1	XP_002524955.1	"ubiquitin-protein ligase, putative [Ricinuscommunis]"	568 bits (1463)	1E-159	379/818 (46%)
Glyma18g07530.1	XP_002279931.1	PREDICTED: hypothetical protein [Vitisvinifera]	495 bits (1272)	1E-137	267/475 (56%)
Glyma18g08220.1	ADC45395.1	HSP90-1 [Glycine max]	1116 bits (2885)	0	583/702 (83%)
Glyma18g08390.1	XP_002266660.1	PREDICTED: hypothetical protein [Vitisvinifera]	668 bits (1723)	8E-190	369/649 (56%)
Glyma18g13290.1	XP_002520371.1	"basic 7S globulin 2 precursor small subunit,putative [Ricinus communis]"	739 bits (1907)	4E-211	363/548 (66%)
Glyma18g41340.1	ACU15822.1	unknown [Glycine max]	415 bits (1065)	6E-114	207/225 (92%)
Glyma18g41350.1	AAG34801.1	glutathione S-transferase GST 11 [Glycinemax]	348 bits (892)	9E-94	179/222 (80%)
Glyma18g41410.1	ACU14762.1	unknown [Glycine max]	451 bits (1160)	3E-125	224/225 (99%)
Glyma18g41820.1	XP_002276211.1	PREDICTED: hypothetical protein [Vitisvinifera]	556 bits (1432)	2E-156	263/315 (83%)
Glyma18g43540.1	AAL08212.1	putative ammonium transporter AMT2 [Lotusjaponicus]	778 bits (2009)	6E-223	382/486 (78%)
Glyma18g44120.1	ACU19112.1	unknown [Glycine max]	399 bits (1024)	4E-109	201/203 (99%)
Glyma18g44630.1	CBI18037.3	unnamed protein product [Vitis vinifera]	749 bits (1933)	4E-214	383/551 (69%)
Glyma18g45260.1	ACU20194.1	unknown [Glycine max]	574 bits (1478)	1E-161	291/327 (88%)
Glyma18g45710.1	XP_002320075.1	predicted protein [Populus trichocarpa]	148 bits (373)	1E-33	76/142 (53%)
Glyma18g46920.1	ABQ88337.1	OAS-TL3 cysteine synthase [Glycine max]	701 bits (1809)	6E-200	359/372 (96%)
Glyma18g47600.1	ACU19761.1	unknown [Glycine max]	672 bits (1732)	6E-191	342/345 (99%)
Glyma18g47720.1	ACU24425.1	unknown [Glycine max]	411 bits (1056)	1E-112	199/200 (99%)
Glyma18g48310.1	ACU15320.1	unknown [Glycine max]	316 bits (808)	3E-84	159/175 (90%)
Glyma18g51750.1	ABF81426.1	NBS-LRR type disease resistance protein [Populustrichocarpa]	273 bits (698)	7E-71	170/456 (37%)
Glyma18g52250.1	ACJ84413.1	unknown [Medicago truncatula]	559 bits (1440)	2E-157	276/315 (87%)
Glyma19g01780.1	O49858.1	RecName: Full=Cytochrome P450 82A3; AltName:Full=Cytochrome P450 CP6	931 bits (2406)	4E-269	455/497 (91%)

Table 2.2. (cont.)

GlymaID	NCBI Tophit 1	Tophit 1 Annotation	Tophit 1 Score	Evalue	Tophit 1 Identity
Glyma19g03010.1	ACU19612.1	unknown [Glycine max]	731 bits (1887)	6E-209	347/453 (76%)
Glyma19g11360.1	XP_002513562.1	"receptor serine/threonine kinase, putative[Ricinus communis]"	481 bits (1236)	2E-133	241/440 (54%)
Glyma19g19680.1	P17928.2	RecName: Full=Calmodulin; Short=CaM	298 bits (761)	1E-78	149/149 (100%)
Glyma19g22590.1	XP_002265308.1	PREDICTED: similar to LHT1 (LYSINE HISTIDINETRANSPORTER 1); amino acid transmembrane transporter [Vitis vinifera]	711 bits (1833)	1E-202	344/441 (78%)
Glyma19g33310.1	ACU24649.1	unknown [Glycine max]	371 bits (952)	6E-101	184/186 (98%)
Glyma19g36400.2	CAA85321.1	protein containing C-terminal RING-finger [Lotusjaponicus]	886 bits (2289)	2E-255	426/546 (78%)
Glyma19g40090.2	AAD03598.1	ethylene response sensor [Vigna radiata]	1175 bits (3039)	0	588/636 (92%)
Glyma19g40560.1	ABC26919.1	WRKY51 [Glycine max]	452 bits (1162)	4E-125	232/287 (80%)
Glyma19g42220.1	AAW78864.1	respiratory burst oxidase 2 [Medicago truncatula]	1484 bits (3841)	0	720/836 (86%)
Glyma19g43460.1	ACU14858.1	unknown [Glycine max]	415 bits (1065)	6E-114	192/204 (94%)
Glyma19g44380.1	ABS18439.1	WRKY43 [Glycine max]	440 bits (1130)	3E-121	216/257 (84%)
Glyma20g11530.1	XP_002525309.1	"receptor protein kinase, putative [Ricinuscommunis]"	448 bits (1152)	9E-124	233/372 (62%)
Glyma20g23080.2	BAF36056.1	calreticulin-1 [Glycine max]	686 bits (1770)	2E-195	330/420 (78%)
Glyma20g26350.1	ACM89596.1	leucine-rich repeat disease resistance protein[Glycine max]	683 bits (1761)	2E-194	348/397 (87%)
Glyma20g26610.1	ACU14269.1	unknown [Glycine max]	467 bits (1200)	1E-129	224/225 (99%)
Glyma20g26940.1	ACD13216.1	zinc finger protein [Cicer arietinum]	255 bits (649)	1E-65	154/288 (53%)
Glyma20g27740.1	ACM89479.1	cysteine-rich protein [Glycine max]	899 bits (2322)	3E-259	455/664 (68%)
Glyma20g31840.3	ACU24200.1	unknown [Glycine max]	416 bits (1067)	8E-114	212/215 (98%)
Glyma20g35180.1	BAA81736.1	GmMYB29B2 [Glycine max]	560 bits (1442)	2E-157	272/272 (100%)
Glyma20g35270.1	ACR39367.1	Aux/IAA protein [Glycine max]	400 bits (1026)	3E-109	201/210 (95%)
Glyma20g35630.1	ACU20151.1	unknown [Glycine max]	717 bits (1850)	9E-205	351/354 (99%)

Table 2.3. Annotations of 11 candidate genes.

GlymaID	Annotation
Glyma01g31750.1	Pfam: Dirigent-like protein Panther: NUCLEOPORIN-RELATED
Glyma07g05480.1	Pfam: Dimerization domain Pfam: O-methyltransferase Panther: O-METHYLTRANSFERASE KOG: Hydroxyindole-O-methyltransferase and related SAM-dependent methyltransferases GO: O-methyltransferase activity GO: O-methyltransferase activity
Glyma10g02370.2	Pfam: ABC transporter Pfam: ABC transporter transmembrane region Panther: ATP-BINDING CASSETTE TRANSPORTER KOG: Multidrug resistance-associated protein/mitoxantrone resistance protein, ABC superfamily GO: transport GO: transmembrane transport
Glyma10g42820.3	Pfam: Arabidopsis phospholipase-like protein (PEARLI 4) Panther: NIPPED-B-LIKE PROTEIN (DELANGIN) SCC2-RELATED
Glyma11g29720.1	Pfam: WRKY DNA -binding domain GO: transcription factor activity GO: sequence-specific DNA binding GO: regulation of transcription, DNA-dependent
Glyma14g00870.1	Pfam: Aldo/keto reductase family Panther: ALDO/KETO REDUCTASE KOG: Voltage-gated shaker-like K+ channel, subunit beta/KCNAB
Glyma17g03340.1	Pfam: Pathogenesis-related protein Bet v I family GO: response to biotic stimulus GO: defense response
Glyma17g03350.2	Pfam: Pathogenesis-related protein Bet v I family GO: response to biotic stimulus GO: defense response
Glyma18g43540.1	Pfam: Ammonium Transporter Family Panther: AMMONIUM TRANSPORTER KOG: Ammonia permease
Glyma18g45260.1	Pfam: NAD dependent epimerase/dehydratase family Panther: NAD DEPENDENT EPIMERASE/DEHYDRATASE KOG: Flavonol reductase/cinnamoyl-CoA reductase GO: catalytic activity GO: cellular metabolism GO: coenzyme binding
Glyma19g42220.1	Pfam: Ferric reductase like transmembrane component Pfam: EF hand Pfam: Ferric reductase NAD binding domain Pfam: Respiratory burst NADPH oxidase Pfam: FAD-binding domain Panther: NADPH OXIDASE

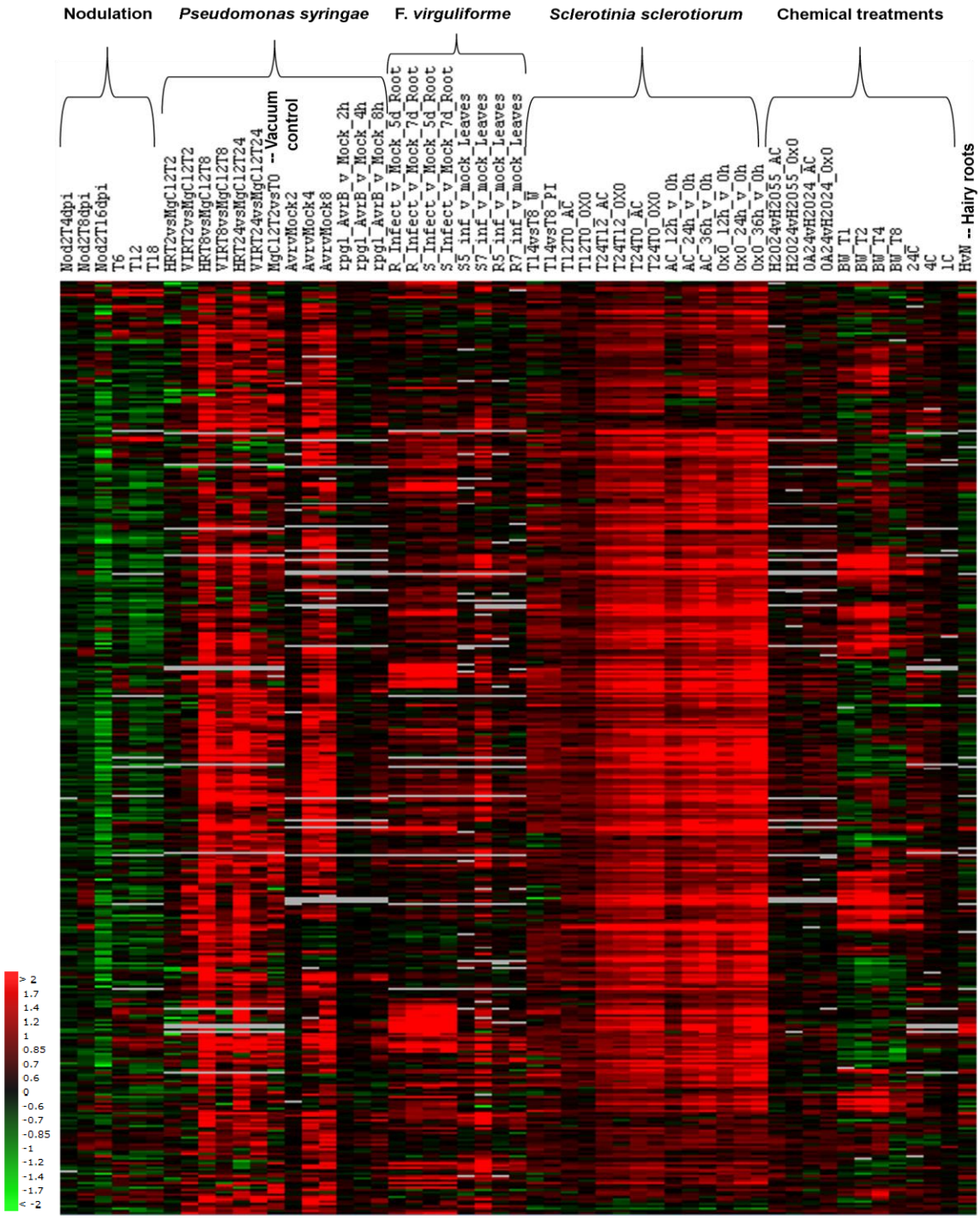


Figure 2.1. Hierarchical clustering of 411 soybean genes differentially expressed. Red indicates increased expression level in response to treatments; green indicates decreased level. The annotations above the figure represent the microarray time courses that measured the changes in transcript abundance in response to treatment with nodulation, pathogens, chemicals and hairy root. 411 genes were identified as being coordinately induced in response to the pathogen treatments (*P. syringae, S. sclerotiorum* and *F. virguliforme*).

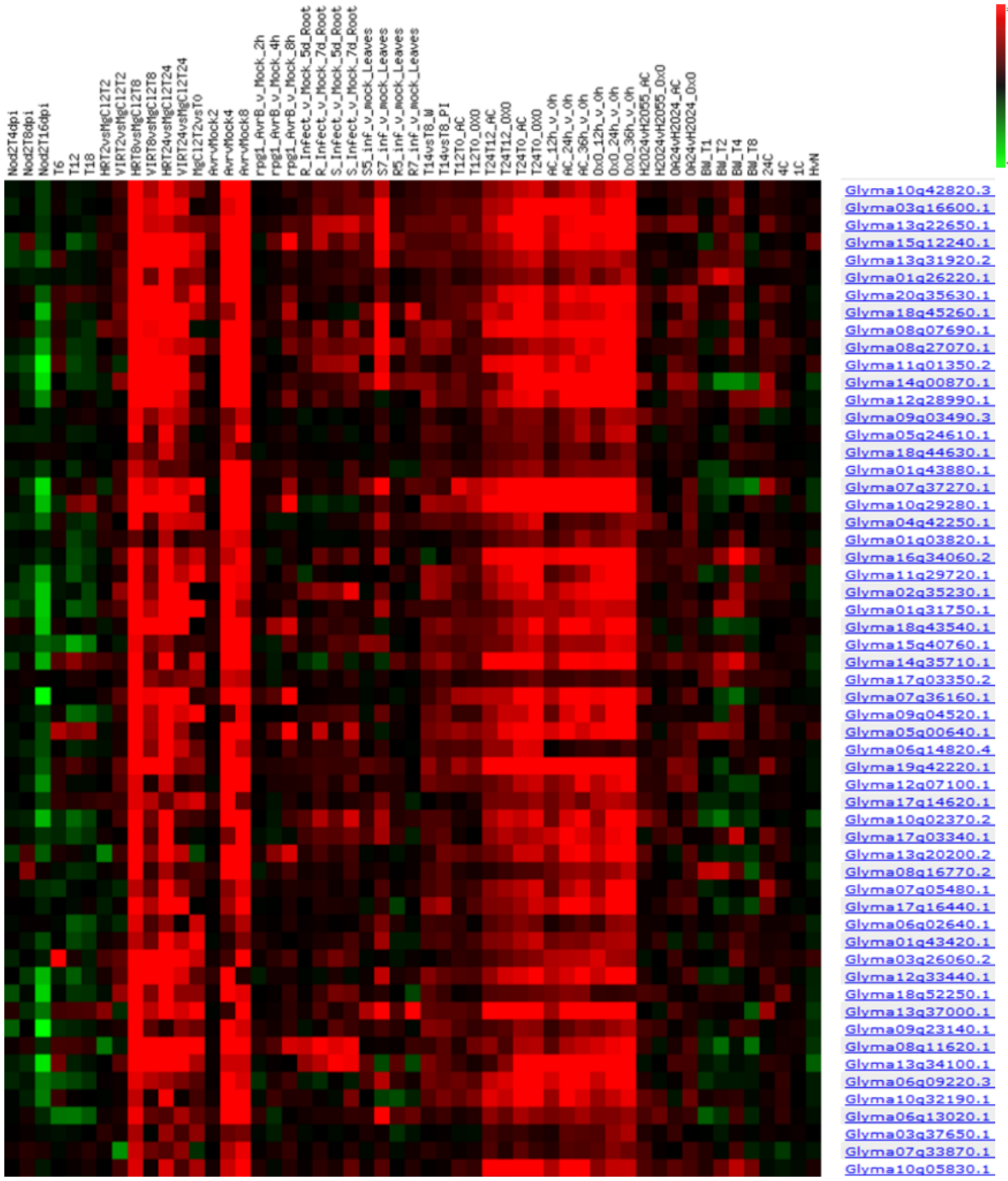


Figure 2.2. Fuzzy k-means clustering of 56 gene candidates in Centroid 3. This centroid was obtained based on the pattern of having genes that were up-regulated in pathogen experiments and an inversed or no expression change displayed in the nodulation studies, and weak or nonexistent change upon treatment with herbicides. Genes were assigned to this centroid using a membership cutoff at 0.2. Genes with missing data in any of the experiments were removed from the clustering.

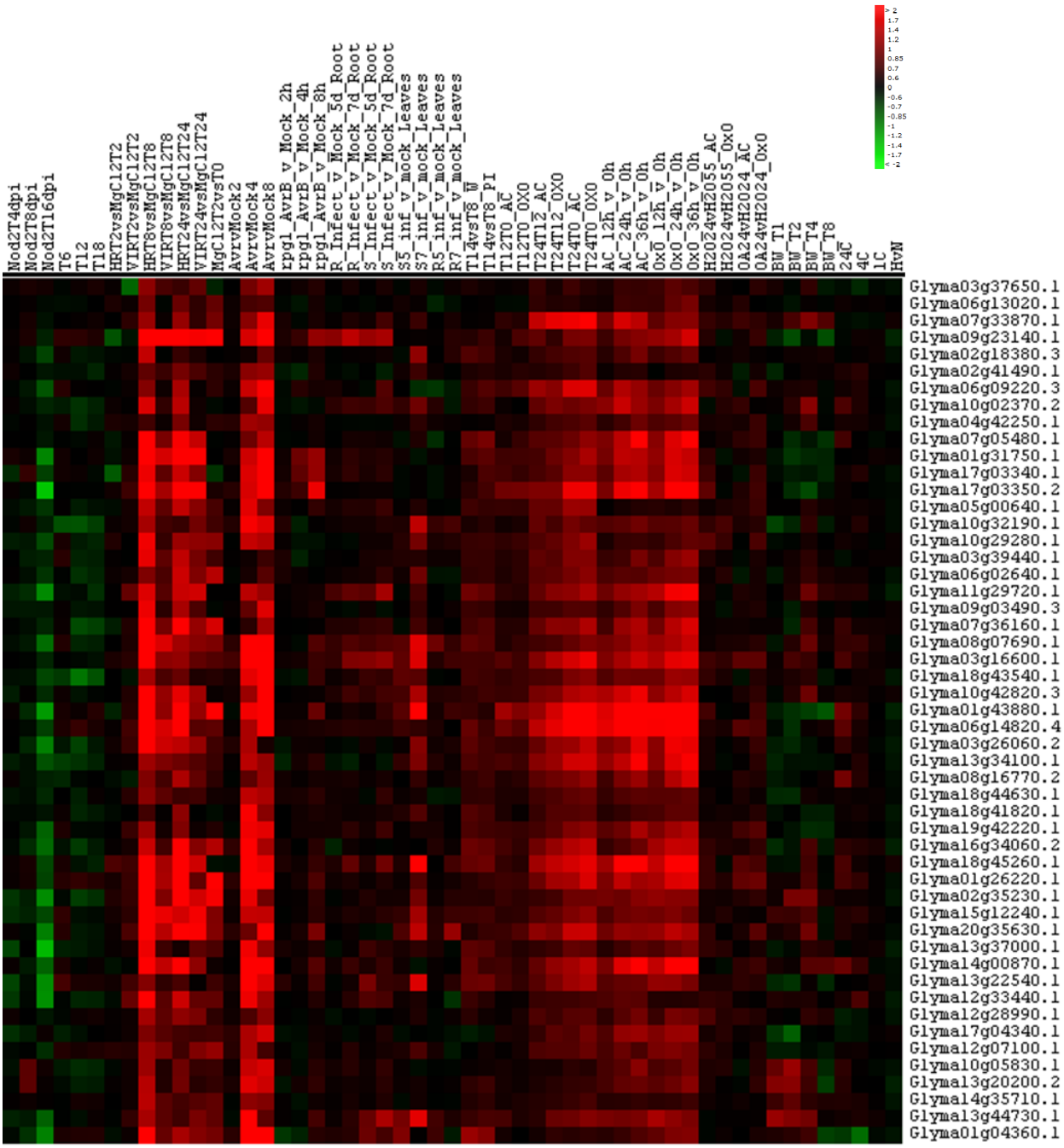


Figure 2.3. Hierarchical cluster of 51 genes differentially expressed. To facilitate down-stream research projects that will look at gene expression in transgenic hairy-root cultures, the candidate list was obtained by removing genes whose expression increased in hairy roots when compared to normal roots.

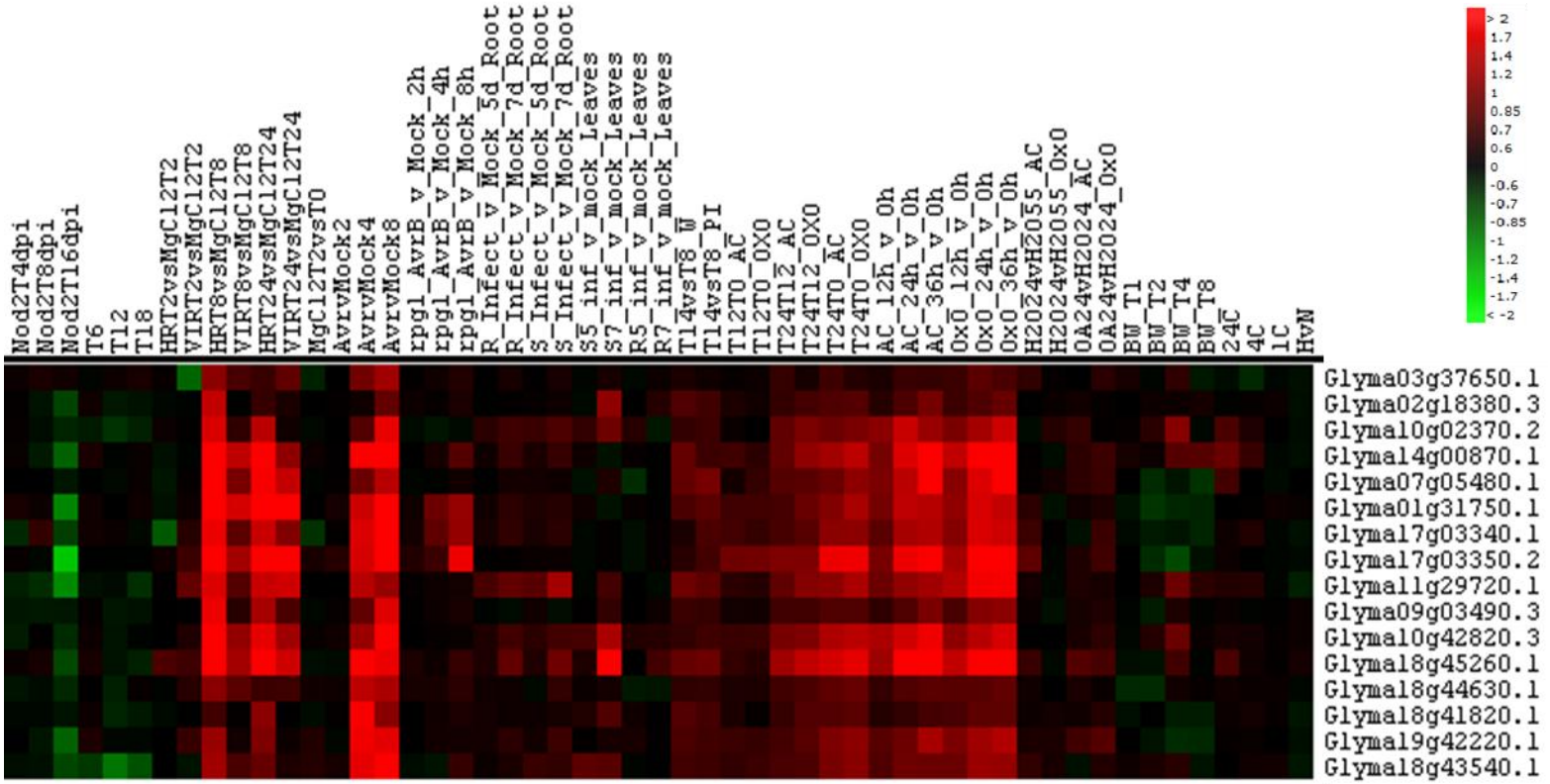


Figure 2.4. Hierarchical cluster of 16 genes differentially expressed. Genes that were induced in leaves by the stress of vacuum infiltrating 10 mM $MgCl_2$ into their apoplastic space were removed from the previous list of 51 genes.

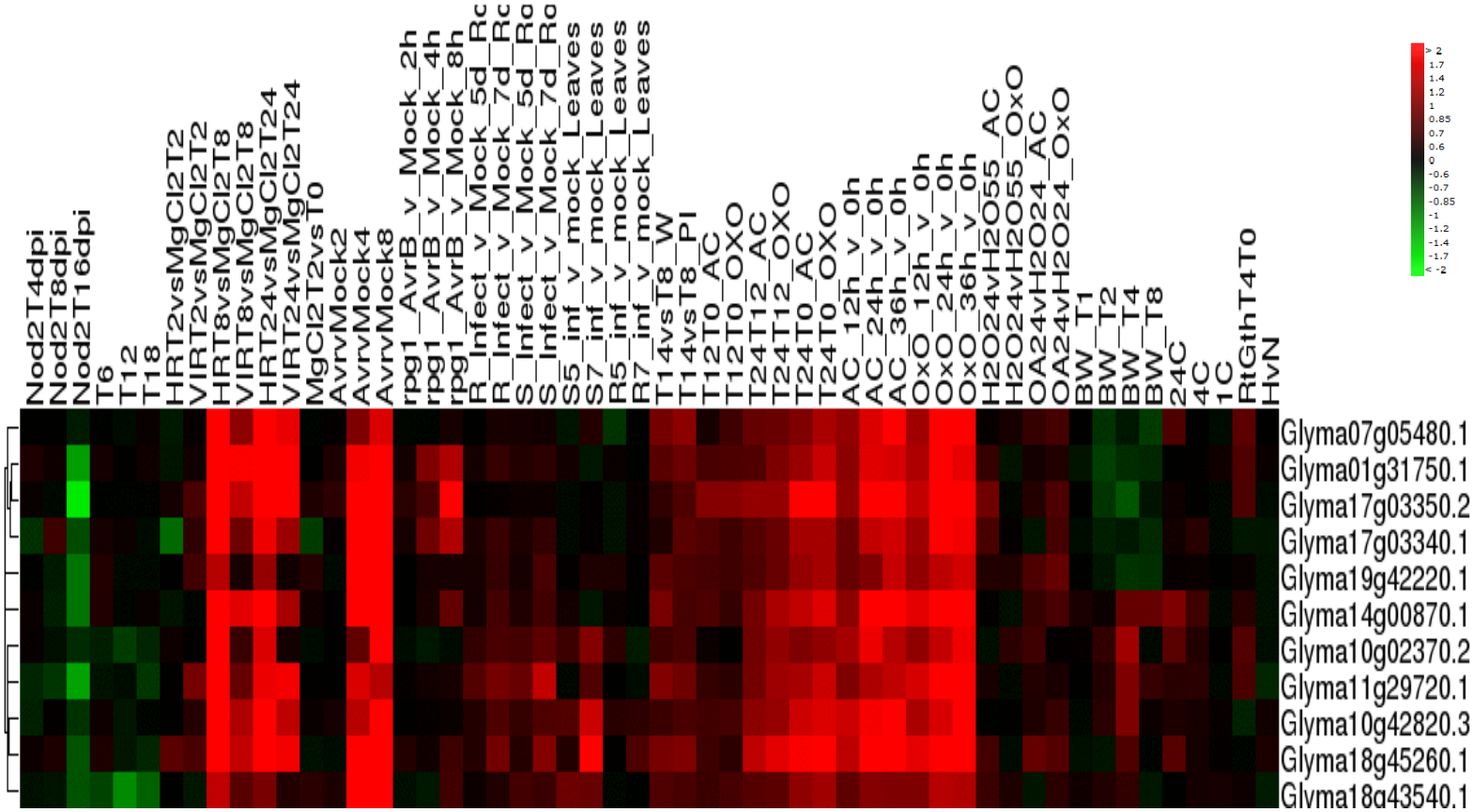


Figure 2.5. Hierarchical cluster of 11 genes of interest. The expression of genes during a root growth experiment was added to verify that the selected genes are not strongly changing during normal root growth. By removing five genes (Glyma02g18380.3, Glyma03g37650.1, Glyma09g03490.3, Glyma18g41820.1, Glyma18g44630.1) that showed very weak expression across all treatments, this final list of 11 genes was induced by pathogens, but not induced by symbiotic *Bradyrhizobium*, 10mM $MgCl_2$ infiltration, or hairy root.

CHAPTER III

CLONING OF SELECT SOYBEAN PATHOGEN-RESPONSIVE GENES FOR OVER-EXPRESSION IN ARABIDOPSIS

I. ABSTRACT

Full-length cDNA of six identified pathogen-specific responsive genes of interest from soybean (*Glycine max*) were cloned into *Escherichia coli*. Two of the six, corresponding to an O-methyltransferase (Glyma07g05480.1) and a dihydroflavonol-4-reductase/cinnamoyl-CoA reductase (DFR/CCR) (Glyma18g45260.1), were cloned into an *Agrobacterium tumefaciens* binary vector pBIN-mgfp5-ER by replacing the *gfp* gene with the cloned cDNA. Following sequence verification of correct cloning, the cloned genes were transformed into *Arabidopsis thaliana* by floral dip and kanamycin selection. Transgenic T1 plants expressing the O-methyltransferase were mostly dwarfed, and because it is uncertain if the reduced size was due to stunting on selective media or due to the transgene, further studies on these plants will be conducted at a later date with T2 plants. Transgenic T1 plants expressing the DFR/CCR were normal in size and therefore a disease study was conducted on these plants by hand-infiltrating a few individual leaves with *Pseudomonas syringae* DC3000 with or without the avirulence factor *avrRpt2*. Leaves were inoculated with approximately $5X10^4$ cells/ml and macerated leaves were dilution plated to determine bacterial propagation during a 4 day period. Comparing the DFR/CCR transgenics to parent plants indicated that this gene did not enhance defense in Arabidopsis and might have made the plants slightly more susceptible.

II. INTRODUCTION

A major goal in plant breeding is to identify sequences that are associated with a trait. One can identify these genetic regions by producing random DNA markers that are polymorphic between parents, and then using mapping methods to determine which of these well-distributed sequence tags are frequently associated with the trait of interest. For diseases such as many viral and bacterial diseases, where single genes can give black-and-white differences between susceptibility and resistance, mapped-based identification of these genetic regions is quite straight forward. However, for resistance that is controlled by multiple genes, where each gene provides only a partial effect, identifying these quantitative trait loci (QTL) can be very challenging and non-traditional gene identification tools may be needed.

High-throughput gene expression analysis is a tool that can be used to identify genes that are affected (expressed or repressed) in response to a given treatment. Hierarchical clustering of 5,417 soybean (*Glycine max*) genes that were shown to be altering in various disease studies (Chapter II), allowed for the identification of 11 pathogen-specific genes that might be involved in providing some defense to diseases controlled by QTL. Verification of a positive effect of these genes on defense in soybean is an ultimate goal, but as transformation of soybean is still very time consuming, one can first determine if a given plant gene provides general defense across plant species, and use the easy to transform (Clough and Bent, 1998) plant *Arabidopsis thaliana* to test gene function.

If one identifies a candidate gene of interest from a crop, and that gene happens to be highly conserved within the plant kingdom, one could study gene function in Arabidopsis by obtaining Arabidopsis mutant of that gene of interest. Several non-profit labs have generated

and made readily available, a vast collection of T-DNA insertion mutants, which provides a high probability of finding an Arabidopsis mutant with a T-DNA inserted into a homolog of a crop gene of interest, with search and ordering on-line from The Arabidopsis Information Resource website (http://www.arabidopsis.org/servlets/Search?action=new_search&type=germplasm). Therefore, if one finds a gene in soybean, such as one of the 11 pathogen induced genes identified previously (Chapter II); one could obtain Arabidopsis mutants of the gene homolog and test for functional complementation of this mutant with the soybean gene, and thus provide evidence of a function of the soybean gene. Such complementary gene function between plant species often works if the genes are well conserved. For example, using an ndr1 Arabidopsis mutant, Cacas et al. (2011) showed that a coffee NDR1 homolog functioned similarly as the well-characterized NDR1, as the coffee gene fully restored *P. syringae* resistance to this Arabidopsis mutant. Soybean is no different in regard to having some of its genes showing similar function in Arabidopsis.

Thanks to Arabidopsis's numerous advantages in genetic research, it has become common practice, when studying a gene of interest from another plant, to introduce this gene into Arabidopsis and overexpress it. By exposing the transformed plants to different conditions (e.g. exposure to pathogens), it is possible to study the gene's function in response to this condition. For instance, Cho et al. (2012) cloned a pathogen-inducible WRKY cDNA from watermelon seedlings shortly after inoculating them with pathogens. Studying target gene expression in transgenic Arabidopsis plants overexpressing this gene pointed to the watermelon WRKY gene's positive regulatory role in plant resistance to pathogens. The success of this experiment suggests that it could be repeated with plants other than watermelon: it

should therefore be possible to clone full-length cDNA of soybean pathogen responsive genes into *Agrobacterium tumefaciens* vectors, and use these constructs to transform Arabidopsis. Transformed plants can then be used to test if the genes of interest enhance disease resistance.

Expressing another crop gene in Arabidopsis is easier than in other plants, thanks to the floral dip protocol (Clough and Bent, 1998). This method is used to transform Arabidopsis by inoculating immature flowers with *A. tumefaciens*. Because it uses intact plant tissues, it avoids unwanted genetic changes such as chromosome breakage. The disarmed Agrobacterium strain GV3101 is used for this protocol: the DNA containing tumor-inducing genes is removed from the Ti plasmid, and an additional plasmid, pBIN-mgfp5-ER (Haseloff et al., 1997) has been added to allow use of the bacterial strain as a vector for plant transformation. The binary plasmid, pBIN-mgfp5-ER (GenBank ID: U87973.1), was derived from pBI121. This plasmid contains kanamycin selectable (*nptII*) gene controlled by the NOS promoter, as well as the endoplasmic reticulum incorporated GFP gene controlled by the cauliflower mosaic virus 35S promoter. Genes of interest, selectable markers or reporter genes (e.g. GUS or GFP) can all be introduced into plants through these vectors: in particular, the modified pBIN-mgfp5-ER contains an *nptII* gene, which confers resistance to the antibiotic kanamycin. As a result, plant cells that have successfully incorporated the T-DNA acquire resistance to kanamycin. Previous studies have confirmed this method for selection of successfully transformed plants (Bent et al., 1994). A map of pBIN-mgfp5-ER is provided in Figure 3.2.

Here we present the cloning of two of the 11 identified soybean pathogen responsive genes (Chapter II), an O-methyltransferase (Glyma07g05480.1) and a dihydroflavonol-4-reductase/cinnamoyl-CoA reductase (DFR/CCR) (Glyma18g45260.1). Phenotypic changes to the

Arabidopsis plants were noted, and the transgenic DFR/CCR plants were assayed for change in defense against the common bacterial pathogen, *P. syringae*.

III. MATERIALS AND METHODS

cDNA synthesis of genes of interest from the soybean genome

As we were interested in cloning genes induced by pathogens, we used total RNA isolated from soybean unifoliates leaves 8 hours after vacuum infiltration inoculation with an incompatible/avirulent strain of *P. syringae* (Zou et al., 2005) as our template for cDNA synthesis. This RNA was previously analyzed in a transcriptomic study (Zou et al., 2005) and found to have a very strong induction of transcript expression changes. RNA was run on a BioAnalyzer RNA chip to verify that it was not degraded. Ambion® DNase I (RNase-free) was used to remove DNA contamination from RNA, followed by a QIAGEN® RNeasy® Mini Kit for RNA cleanup. Amplification of total cDNA was generated using the Invitrogen™ SuperScript™ III First-Strand Synthesis System on 1.665ug DNA-free RNA and oligo (dT) as the polymerization primer.

RT-PCR cloning

Primers were designed to select for each of the 11 genes identified in Chapter II. Primer length was 18-22 base pair, which is long enough for provide specificity and short enough for primers to bind easily to the template at the annealing temperature. Primer sequences were Blast'd against the soybean genome to determine specificity. Primer pair oligonucleotides selected and used, and their product length was described in Table 3.1. The target genes were amplified via 40 PCR cycles at 95°C for 30 seconds, 55°C for 30 seconds, and 68°C for 2 minutes, followed by a final extension of 2 minutes at 68°C. PCR products were analyzed on 1% agarose

gel to identify reactions that produced only a single PCR product. Reactions that produced no or multiple amplicons were repeated in efforts to obtain a pure product.

Clone RT-PCR product into plasmid vector

For reactions that produced a single pure band on a gel, the cDNA fragments were cloned into a vector using the QIAGEN® PCR Cloning Kit. The pDrive Cloning vector supplied with this kit allows blue-white colony screening, and it has built-in ampicillin and kanamycin regions in the plasmid for antibiotic selection. Also, this vector contains multiple unique restriction sites, making it easy for restriction enzyme cutting of recombinant plasmids. The vector is constructed with standard sequencing primers T7 and SP6 on each side of cloning sites. A map of the pDrive Cloning Vector is provided in Figure 3.1. The pDrive Cloning Vector provided a U residue and was easily connected with the single A overhang of the PCR products by *Taq* DNA polymerases. After 2 hours of UA-based ligation at 4°C, ligation mix was transformed into *Escherichia coli* (NEB® Turbo Competent cells). Cells were thawed on ice and DNA was added immediately when the last ice crystal disappeared. To increase transformation efficiency, cells and the ligation mix were incubated together on ice for 30 minutes without vortex disturbance followed by a 42°C heat-shock for exactly 30 seconds. Subsequently, the mixture was incubated in 950 µl SOC medium (2% Bacto™ Tryptone, 0.5% Bacto™ yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose, water) with vigorous shaking at 37°C for 1 hour and spread on warm kanamycin selection Luria Bertani (LB) plates (10 g of Bacto™ Tryptone, 5 g of Bacto™ yeast extract, 5 g of sodium chloride, 15 g agar per liter) with 80 µl of X-gal at 20 mg/ml and 80 µl of IPTG at 24 mg/ml on the surface. Allow overnight colony formation at 37°C incubator. To determine the correct orientation of

insertion, PCR was conducted on the pDrive clones using a gene-specific primer and a vector associated T7 primer. Once the right insertion orientation was confirmed, the purified plasmid (Qiagen) was sequenced to ensure the correct and full-length gene was cloned, and then the plasmid was digested with *Bam*HI-HF and *Sac*I-HF in buffer 4 (both enzymes and buffer from NEB®) for 2 hours at 37°C. In a separate digestion, the Agrobacterium binary vector pBIN-mgfp5-ER (Haseloff et al, 1997) was digested with the same enzyme to release the *gfp* gene, allowing it to be replaced with our gene of interest, after gel purification.

Takara® RECOCHIP is designed to recover DNA fragments from an agarose gel. After electrophoresis, DNA bands corresponding to the inserts and to the pBin-mgfp5-ER minus the gfp gene were recovered following manufacture protocol (RECOCHIP). Since long exposure to UV light may damage DNA, the strong UV light in observation chamber was not recommended, and a hand-held 100W long-wavelength UV lamp was used to visualize the bands. When located the bands of interest, the RECOCHIP was inserted into the gel with its non-woven polyester facing the band. Then the electrophoresis was performed again, so that the band would run into the space between the non-woven fabric and its attached cellulose dialysis membrane. Electrophoresis was stopped after 10 minutes and the DNA recovered by placing the RECOCHIP in a collecting tube and centrifuged. After recovery, the gel was observed under a UV illuminator again to confirm that the correct bands were recovered. Quick® T4 DNA ligase was used to combine separated cut binary vector with our insert.

Ligation mixtures were electroporated into electrocompetent cells of *A. tumefaciens* strain GV3101, using the 'Agrobacterium' preset on the BIO-RAD MicroPulser™ Electroporator. Electrocompetent cells, 80 µl, were thawed on ice, and then 1 µl of the ligation mix was added,

and pipetted into a chilled 0.2 cm electroporation cuvette. A single 2.5kV electrical pulse to the cuvette was performed. Following the electroporation, 1 ml cold MGL medium (5 g of Bacto™ Tryptone, 2.5 g of Bacto™ yeast extract, 2.7 g of NaCl, 10 g of mannitol, 2.32 g of sodium glutamate, 0.5 g of $K_2HPO_4 \cdot 3H_2O$, 0.2 g of $MgSO_4$, and 2µg of biotin per liter) was added to the cuvette to remove cells, and incubated with shaking for 2 hours, and then spread onto kanamycin selection plates and incubated at 25°C for three days to obtain transformed colonies. Single colony from the transformed colonies was picked to streak out on a fresh plate. M13 primers (forward, 5'-GTAAAACGACGGCCAGT-3' and reverse 5'- AACAGCTATGACCATG-3') were used in PCR of this colony to verify Agrobacterium had the correct plasmid.

Plant growth and transformation into Arabidopsis

Nine *A. thaliana* Columbia (Col-0) seeds were distributed evenly on the surface of 3.5 inch pots containing LG Sunshine Mix potting soil, covered with a plastic dome, and stored in a cold room. After three days, plants were moved to a growth chamber at 23°C in a 12/12 (day/night) photoperiod for 15 days with a light intensity of ~180 µmol photons $m^{-2}s^{-1}$. The pots were randomly placed in the tray inside of the growth chamber and allowed to grow until flowering.

Agrobacterium cultures with the verified gene cloned into the binary vector, were streak on LB medium, supplemented with kanamycin at 50 µg/ml. After three days grown in room temperature, an isolated single colony was used to inoculate 5 mL liquid LB and the culture was incubated at 28°C with shaking at 250 rpm for about 18 hours. A 5 ml aliquot of the culture was used to seed a new liquid 50 ml LB culture which was allowed to grow to a turbid culture by

vigorous agitation overnight at approximately 25°C. The culture was then pelleting the next day and resuspending them in 5% sucrose solution and surfactant Silwet L77 at 0.03%.

Plants at the correct flowering stage (Clough and Bent, 1998) were inverted into the Agrobacterium suspension such that the Arabidopsis inflorescences were fully submerged into the inoculum for 5 seconds. Then plants were placed in a covered flat in a low light area overnight, and moved to growth chamber the next day and cared for until seed maturation. Plants were not watered after they started senescing, and seed harvested once the plants were completed brown and dry.

Seeds were harvested and disinfected with 95% ethanol followed by 5 minutes in 50% bleach – 0.05% Tween-20 and three times rinses in sterile water, and plated on Murashige and Skoog Basal Salt Medium (Sigma) containing 0.8% tissue-culture tested agar (Phyto Technology Laboratory) and kanamycin at 50 µg/ml. Spread approximately 3000 seeds on each plate. The selection plates were sealed with Parafilm® and placed at 4°C refrigerator for three days to maximize the germination. After this vernalization period, plates were moved to growth chamber at 22°C, under 12 hours light periods for two weeks. After the incubation, the green transformed seedlings were selected and transferred carefully to 1.5 inch square pots containing LG Sunshine Mix soil (Figure 3.5).

Pathogen inoculation and sampling

Bacterial inocula consisted of the strain *P. syringae* pv. *tomato* DC3000 with or without the avirulence gene *avrRPT2* suspended in 10 mM $MgCl_2$. Cells were scraped from overnight cultures and dispersed in 25 ml 10 mM $MgCl_2$ to an OD_{600}=0.05 (CO8000 Cell density meter) which corresponds to approximately 5 x 10^7 colony-forming units/ml. After a 1000 fold dilution,

5 x 10^4 colony-forming/ml was obtained. 1cc syringes without needles were used to infiltrate the bacterial inocula solution into leaves of transgenic Arabidopsis plant (with transgene DFR/CCR Glyma18g45260.1). The target leaves were marked with one black dash to indicate inoculated with DC3000 carrying *avrRPT2* and two red markers to indicate inoculated with DC3000 without *avrRPT2*. Time points for tissue harvest and dilution plating were set at 0 h, 48 h, and 96 h post inoculation.

For the T0 set, when the leaves were dry enough after the inoculation (about 2 hours) and showed no color difference (due to water-soaked apoplast) with the non-inoculated ones, the leaves were cut at the petiole with surgical scissors and immediately weigh on balance and weight recorded. Leaves were ground in 200 µl ice cold sterilized water in 1.5 ml microfuge tubes until no visible particles could be seen. Macerated extract was plated on King's B medium and placed at room temperature. Colony counts were observed starting around 1.5 days after plating, and were made as soon as colonies were large enough to be visually recognized. For the T48 and T96 timepoints, leaves were treated similarly. Resuspended bacteria were plated in sequential serial 1:10 dilutions to ensure colonies were countable and determine changes in population levels as described (Hinsch and Staskawicz, 1996).

RNA extraction and cDNA synthesis

The transgenic Arabidopsis leaves (DFR/CCR Glyma18g45260.1) with no pathogen-inoculation and Col-0 wild-type control leaves were used for RNA extraction and expression analysis. After cutting off the leaf with surgical scissors, the leaves were immediately placed into 1.5 ml centrifuge tubes and frozen in liquid nitrogen within 10 seconds. Control leaves were harvested from Arabidopsis Col-0 plants that were not infiltrated as "untouched" control.

Samples were transported under liquid nitrogen in the laboratory and stored at -80°C before RNA extraction, or being immediately performed with RNA extraction.

Total RNA was isolated using QIAGEN® RNeasy® Plant Mini Kit following manufacturer's protocol. RNA samples were quantified by spectrophotometry. RNA quality was determined by results combination of Agilent's BioAnalyzer RNA 6000 Nano Assay (Agilent Technologies, Palo Alto, CA, U.S.A.) and NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, U.S.A.). High quality RNA was reverse transcribed into cDNA with Invitrogen™ SuperScript™ III First-Strand Synthesis System for RT-PCR. The cDNA was then used as the template for PCR amplification.

Semiquantative RT-PCR was performed using Glyma18g45260.1 gene specific primers (forward, 5'-GCCACAGCAAGAGTGTTCAA-3' and reverse 5'-ATCATTGCCAGGGATCTGAA-3'). Ubiquitin primers (forward, 5'-GGCCTTGTATAATCCCTGATGAATAAG-3' and reverse 5'-AAAGAGATAACAGGAACGGAAACATAGT-3') were used as a standard. PCR products were analyzed on 1% agarose gel.

T2 transgenic Arabidopsis

Harvested T2 seeds of transgenic *Arabidopsis thaliana* (DFR/CCR Glyma18g45260.1) were sterilized by chlorine gas for 6 hours by placing seeds in opened microfuge tubes within a bell jar in which a beaker with 100 ml bleach, and then adding 3.5 ml 12N HCl and quickly sealing the chamber. Approximately 20 seeds per plant were spread on the kanamycin selection plate to determine if plants contained the T-DNA.

IV. RESULTS

Full-length cDNA of six candidate pathogen responsive genes of interest from soybean were cloned into *E. coli*. The cDNA correspond to Glyma01g31750.1 (dirigent-like protein), Glyma07g05480.1 (O-methyltransferase, OMT), Glyma10g42820.3 (phospholipase-like protein), Glyma14g00870.1 (aldo/keto reductase family protein), Glyma18g43540.1 (ammonium transporter family protein), and Glyma18g45260.1 (dihydroflavonol-4-reductase/cinnamoyl-CoA reductase protein, DFR/CCR). The process of the 11 select genes is displayed in Table 3.2. We inoculated immature flowers of Arabidopsis plants with Agrobacterium strain GV3101 containing the cDNA cloned into plasmid pBIN-mgfp5-ER, swapping the *gfp* for the cDNA. Insertion of the T-DNA into the Arabidopsis genome resulted in the production of transgenic seeds that grew into transgenic plants, selectable via the T-DNA encoded *nptII* gene.

The OMT (Glyma07g05480.1) and the DFR/CCR (Glyma18g45260.1) were successfully transformed into Arabidopsis. Approximately 100 putative OMT transgenic plants and 51 putative DFR/CCR transgenic plants grew on kanamycin selection plates. After transferring to soil and allowing to grow another 2 weeks, some phenotypic changes were noted in the OMT transgenics, whereas the DFR/CCR transgenics looked normal. Among the 70 OMT transgenic plants, we noted only five had a size comparable with wild type plants, while 54 looked severely dwarfed and the remaining 11 had an intermediate size (Figure 3.6). Another interesting observation was that the transgenic OMT dwarf plants produced rather large-sized inflorescence, similar to that expected of a wild-type parent plant (Figure 3.7 A). Therefore, the leaf rosette size was not proportional with the size of stem and inflorescence, with rosettes averaging about 3 cm in diameter producing inflorescences from 20-30 cm. Additionally, some

OMT transgenic plants demonstrated leaf bleaching or a chlorotic appearance (Figure 3.7 B). All these phenotypic observations for the OMT transgenics suggest that the gene is affecting the overall health and/or development of Arabidopsis. As these abnormal phenotypes could be the result of a negative effect from the selection plates, these plants were left to produce a T2 generation before functional assays will be conducted on them.

The T1 generation of the DFR/CCR transgenics grew to be healthy, without any obvious variations from parental Col-0 plants. These 51 putative transgenics were biased in their numbering, with the largest and healthiest looking plantlets on the selection plates being numbered and transplanted first. Therefore, number 1 was the healthiest, and number 51 was the weakest and most questionable as to whether it was a true transgenic or an 'escape'.

Interactions between Arabidopsis and *P. syringae* have been studied as a model system for plant-pathogen molecular interactions since 1991 (Whalen et al., 1991). This study specifically observed interactions between Arabidopsis and the bacterial pathogen *P. syringae* pv. *tomato* (*Pst*). *Pst* strains were either virulent or avirulent when presented to specific Arabidopsis ecotypes, such that individual ecotypes were resistant to some *Pst* strains and susceptible to others. Plant resistance to pathogens is often controlled by the interaction of single plant resistance genes and single pathogen avirulence genes. The Arabidopsis ecotype Col-0 was found to be susceptible to the *Pst* strain DC3000, but resistant to the same strain when it carried *avrRpt2*: this indicates that a single locus, RPT2, in Col-0 determines resistance.

Soybean DFR/CCR Arabidopsis transgenics were challenged with *Pst* DC3000 carrying the avirulence gene *avrRPT2* to induce the incompatible (or HR, hypersensitive response) and without *avrRPT2* to produce a compatible virulent plant-microbe interactions. These

preliminary results taken from T1 plants suggested that the DFR/CCR transgene (Glyma18g45260.1) has no effect on resistance or, in anything, a minor enhancement of susceptibility.

At both the 48 and 96 h post inoculation (hpi) time points, T1 DFR/CCR leaves infected with *Pst* with *avrRpt2* responded as to be expected with Arabidopsis is infected with a dilute suspension of an avirulent pathogen—no disease symptoms were observed. Likewise, leaves infected with *Pst* lacking *avrRpt2* appeared to be diseased, with the leaves turning yellow and some necrosis along the edges.

Bacterial content in the pulverized plant material was determined for both control and transgenic plants at 0, 48 and 96 hpi (Figure 3.10). Not surprisingly, the number of bacterial colonies increased over time in plants, whether inoculated with the virulent or avirulent strain, but the avirulent strain did not reach as high a population, with about 10-fold lower population at 96 hpi. It was interesting to note that the growth of the bacterial population appeared to be slightly greater in the DFR/CCR transgenic plants compared to the control, suggesting that the transgene might have enhanced susceptibility, which be explained if this enzyme, putatively within the phenylpropanoid pathway, diverted substrate away from an branch leading to defensive phytoalexins. Alternatively, this slight increased growth of the pathogens in the DFR/CCR transgenics might be an artifact of using T1 plants that were not at the exact same growth stage as the Col-0 controls, as the control plants did not go through agar plates and transplantation, and therefore were slightly more advanced. Perhaps, the Col-0 controls were already shunting carbon and nutrients away from leaves to begin the production of inflorescence and therefore, less carbon and nutrients were available for maximal pathogen

population growth, leading to the slight, but noticeable differences in bacterial increase. Transgenic plants were one week younger than the control plants when they were inoculated: it is therefore possible that age difference, rather than defensive capability, could explain the difference in bacterial development between transgenic and control plants: younger plants likely have greater photosynthetic rates, allowing them to produce more photosynthate, which could in turn support a larger bacterial population. The experiment will be repeated with T2 plants to see if the results are consistent.

Semi-quantitative RT-PCR was then performed on the T1 DFR/CCR transgenics to verify expression of the *DFR/CCR* as well as to determine relative expression levels between transgenics. Due to the small leaf size of Arabidopsis, the RNA amounts were very limited for each plant. RNA quality was determined and appeared to be high as indicated by the presence the ribosomal RNA bands (Figure 3.11). Based on equal loading of RNA and ubiquitin standard control, different levels of expression in each plant were observed (Figure 3.12). We designed replication for each plant sample and the expression levels across duplicates were very consistent. There are several explanations on why the expression levels could vary. Because T-DNA inserts randomly, expression variation is due to genome insertion location. The T-DNA could have inactivated a gene that somehow affects transcription, or it could have inserted close to a promoter, and that native promoter can affect transgene expression. Additionally, the region of insertion might be in an area of the genome that is generally weakly expressed, likely due to DNA being tightly wrapped around histones.

T2 seed from the 51 T1 DFR/CCR transgenic plants were assayed for percent kanamycin resistance as a means of verifying that the plants were indeed transformed, and secondly to

ensure that transformed plants would be planted for T3 seed production to allow eventual identification of homozygous transformed lines. Approximately 20 seeds were plated for each plant, and young seedlings were observed for overall health and root development into the agar. Often, dead plants were white and obviously dead, but often some were just yellowish, but growing ok. After another week of growth, these yellowish plants died as well. Plantlets that were green but did not produce a deep root system were considered to be escapes, and not true transgenics and these too died eventually. The most obvious difference is that the real transformants stretch their roots deep down to the plate to get enough nutrients, while the escapes have their roots comparably on the surface. The escapes were eliminated when calculating the survival rate.

According to Mendelian segregation, after the first generation of Arabidopsis self-pollination, three possible genotypes could be found for each inheritable trait: AA, Aa, and aa, with dominant A representing the DRF/CCR transgene. Therefore, a 3:1 ratio was expected for plant survival on kanamycin selection medium. However, the seed counting result showed that the rate was 54% for the 1125 seeds planted, which is close to a ratio of 1:1. Actually the result was not unusual. The reason behind this can be explained if homozygous DFR/CCR was toxic as is often the case when a foreign gene is expressed too strongly in a new host. The DFR/CCR is driven off the CaMV35S promoter, which is highly expressed in Arabidopsis. This high expression off the CaMV35 promoter in the homozygous seedlings might result in weak plants that do not survive on the selection plates. The T3 seed will be plated on kanamycin as well, to verify this hypothesis or see if homozygous T-DNA lines are possible. If the homozygotes are

lethal, then it will be necessary to plate seed on kanamycin selection before each experiment to verify presence of the transgene before moving to soil.

According to our labeling system, we numbered the transgenics from 1 to 51, according to decreasing of their robustness. Plant #45 could be an example of first generation escapes due to the fact that it was weak when picked up for transplanting. However, we were surprised that plant #4 did not survive the selection plates as it appeared to be one of the stronger T1 plants growing on the kanamycin plates. Perhaps this plant was an escape as well, or it died due to the nature of the mutation or expression of the T2 generation.

Attempts to clone and transform candidates from the 11 newly found pathogen inducible genes into Arabidopsis, and eventually soybean, will continue in the near future. Should disease resistance in soybean be effectively increased, the gene sequences can be developed into molecular markers, which can then be used by breeders in the development of resistant varieties, or in the development of transgenics expressing improved resistance. Upcoming work on this research project will provide genes for promoter analysis and identify genes involved in pathogen-specific responses. These promoters will be key tools for studying precision expression of important genes in crops.

Tables and Figures

GlymaID	Forward Primer	Tm(F)-IDT	Reverse Primer	Tm(R)-IDT	Product Size
Glyma01g31750.1	TCAACTCTCCCAGCAAAGGT	56.3	ACCACATGCCCTCAAAAGAG	55.3	828
Glyma07g05480.1	GCTGCTTTGTTTGGTACATTCA	54.2	TTTCAAGCAACAAGCAAGGAT	53.4	1209
Glyma10g02370.2	TGGTAGGCGGGTATAGTTGC	56.7	CTGCCCCACACTCCAGTTAT	56.9	4888
Glyma10g42820.3	GCAGGGGAGACTCAGAAAGA	56.3	TGCATGTGAGTGTGTGCAAT	55.5	1410
Glyma11g29720.1	CACCTATCTCATCAATTT	43.4	AACCAACCCCTTCCAAGA	53.7	1731
Glyma14g00870.1	AAACTCCTCTGCCCAACAGA	56.3	AGAAGTGGCATGGCATGAAT	54.9	1009
Glyma17g03340.1	AACTCACACCCTAAGAGAGCA	55.4	TTCTCATGAAGCAAAGTCAAAC	51.6	648
Glyma17g03350.2	TCTTCCATTTCCATTAAAAAG	46.4	ATTGAATAACACTCCACACAAG	51.0	595
Glyma18g43540.1	AACTTTCATCGGCCACAAAT	53.0	ATGTGTGAGCAAAGCACAGC	56.5	1620
Glyma18g45260.1	GCCACAGCAAGAGTGTTCAA	55.9	ATCATTGCCAGGGATCTGAA	53.9	1087
Glyma19g42220.1	CATCCCCCGTCAACTCCA	57.1	ATTTCCCTCAACTACATAC	46.2	3089

Table 3.1. Forward and reverse primers used for 11 gene candidates.

GlymaID	RT-PCR	pDrive	pBIN	T1 transgenics	Pathogen inoculation	T2 transgenics
Glyma01g31750.1	✓	✓				
Glyma07g05480.1	✓	✓	✓	✓		
Glyma10g02370.2						
Glyma10g42820.3	✓	✓				
Glyma11g29720.1						
Glyma14g00870.1	✓	✓				
Glyma17g03340.1	✓					
Glyma17g03350.2						
Glyma18g43540.1	✓	✓				
Glyma18g45260.1	✓	✓	✓	✓	✓	✓
Glyma19g42220.1						

Table 3.2. Progress of 11 gene candidates.

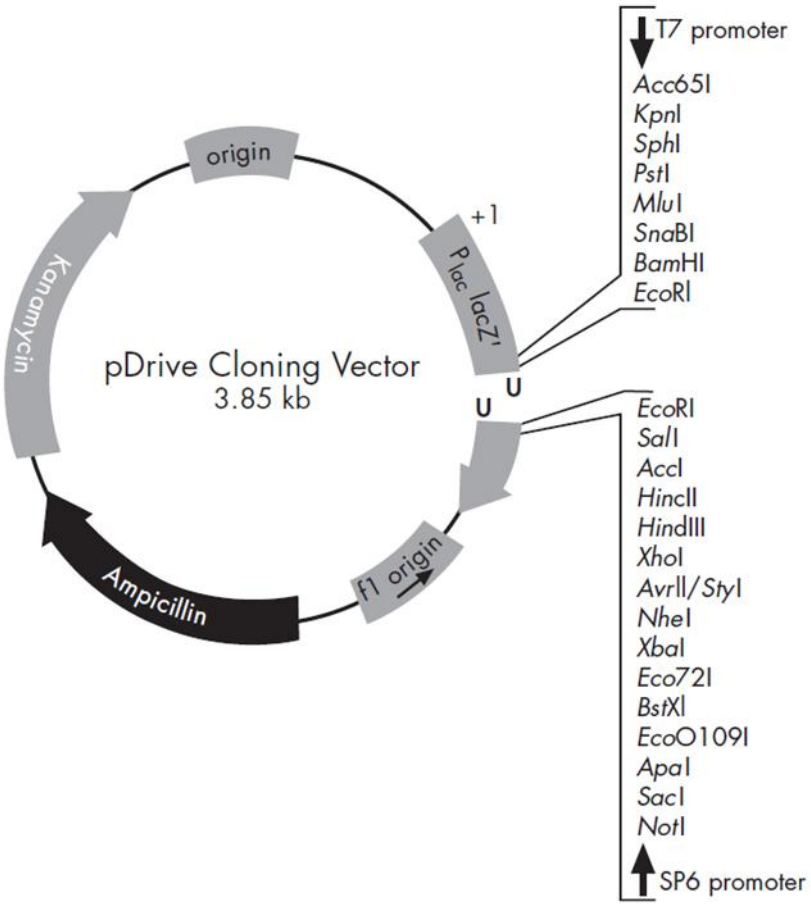


Figure 3.1. Map of pDrive Cloning Vector of QIAGEN® PCR Cloning Kit.

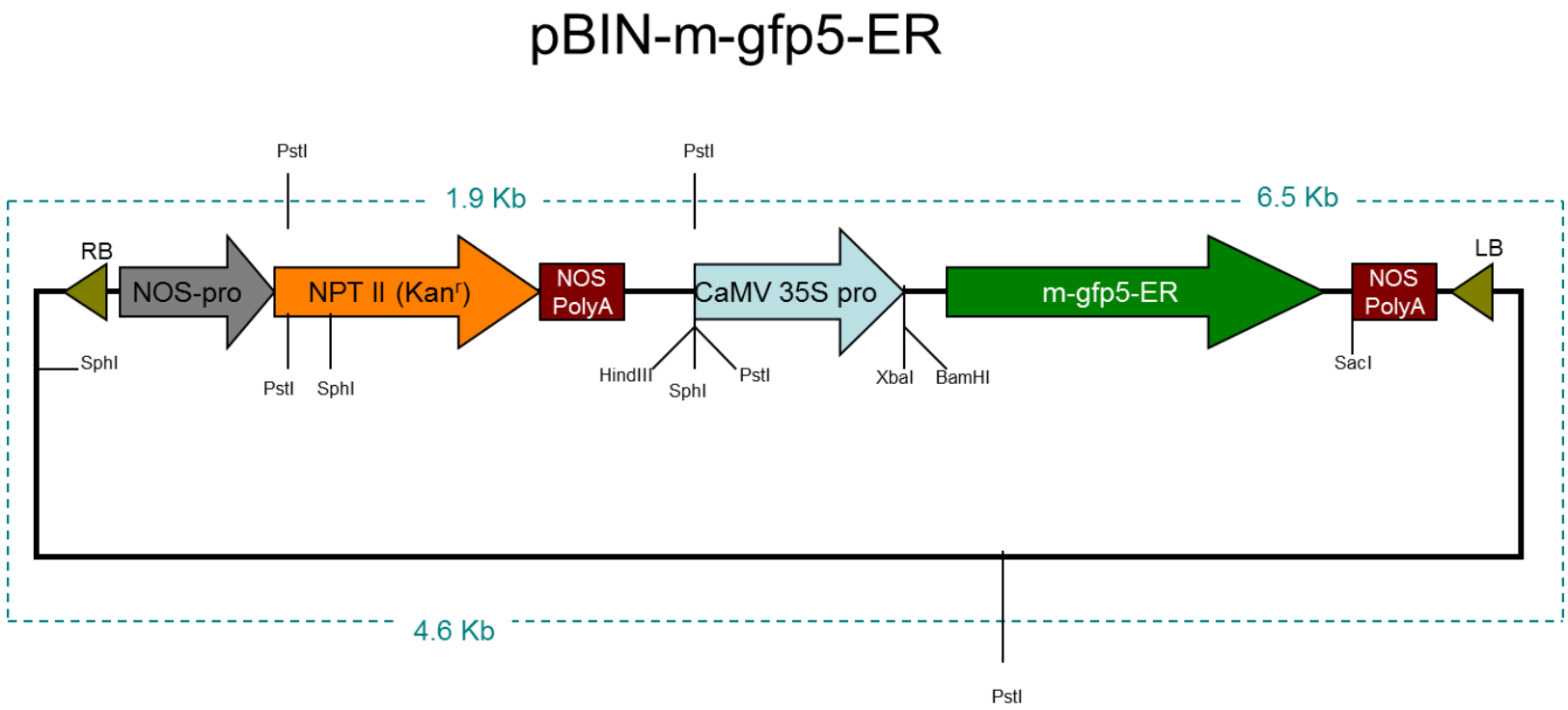


Figure 3.2. Map of pBIN-m-gfp5-ER.

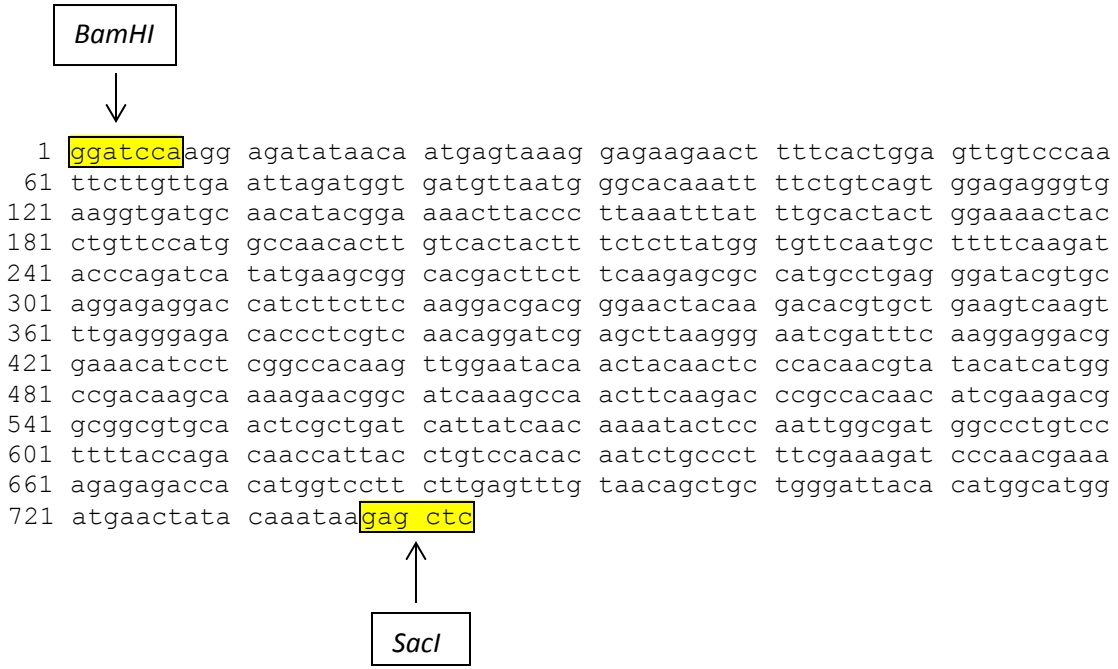


Figure 3.3. Sequence of m-gfp5-ER within the pBIN-m-gfp5-ER. With *BamHI* and *SacI* on each end, the overall length of this gfp is 743 base pair.

Figure 3.4. Seedlings of transformed Arabidopsis with transgene DFR/CCR (Glyma18g45260.1) on the kanamycin selective plates. Transformed plants are easy to identify after about two weeks growth on the selective medium. Transformed seedlings will be green and healthy, whereas the nontransformants will be chlorotic, stunted, and dying. The root system of a transformant will also be much longer than that of the nontransformants.

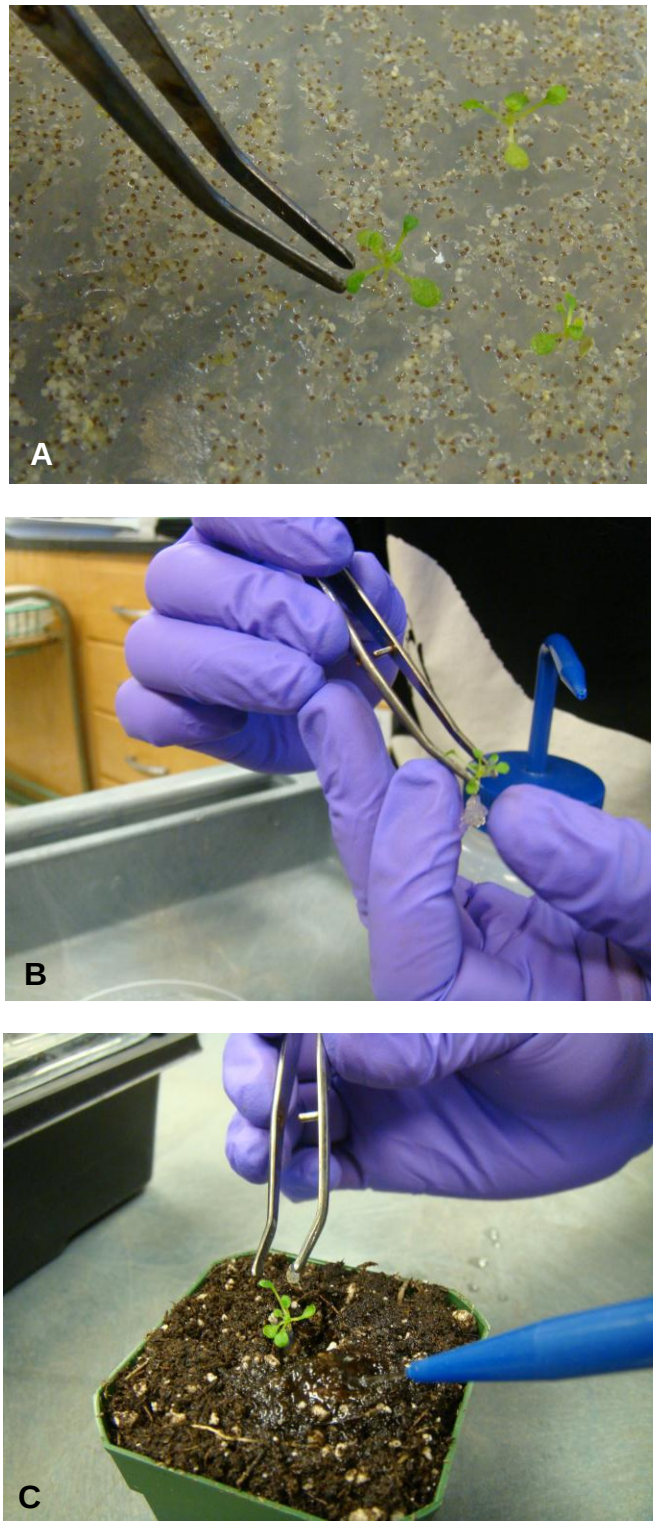


Figure 3.5. Transplant the transformants to soil to allow plants to reach maturation.

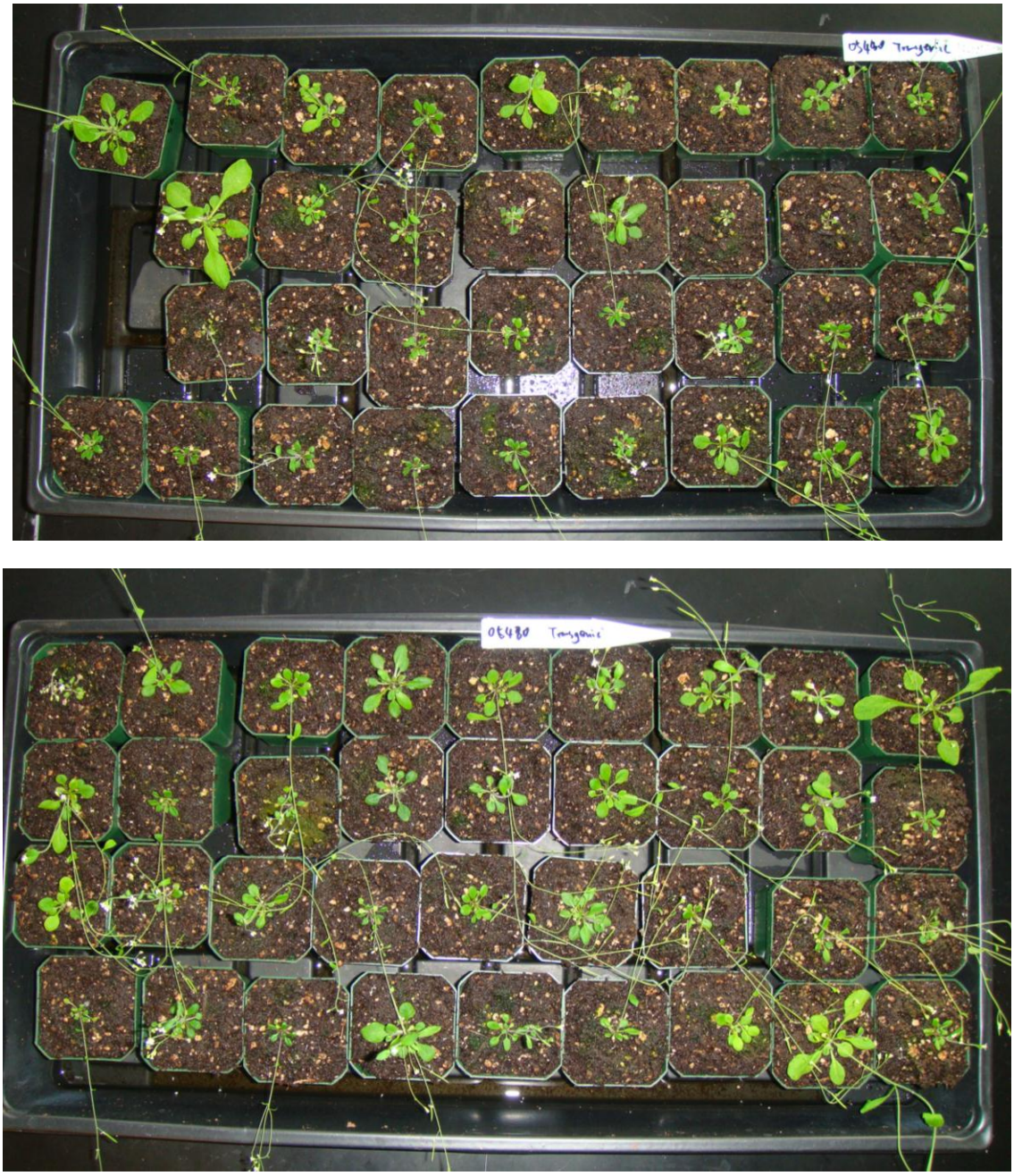


Figure 3.6. 70 mini-pots of transgenic plants of OMT (Glyma07g05480.1). Among the 70 OMT transgenic plants, only five had a size comparable with wild type plants, while 54 looked severely dwarfed and the remaining 11 had an intermediate size.

A

B

Figure 3.7. Select phenotypes of first generation of transgenic plants of OMT (Glyma07g05480.1). A: The transgenic OMT dwarf plants compared with a regular sized transgenic OMT Arabidopsis. B: An example OMT transgenic plant demonstrated leaf bleaching or a chlorotic appearance.

Figure 3.8. Phenotypes of first generation transgenics of DFR/CCR (Glyma18g45260.1).

Figure 3.8. (cont.)

Figure 3.8. (cont.)

Figure 3.8. (cont.)

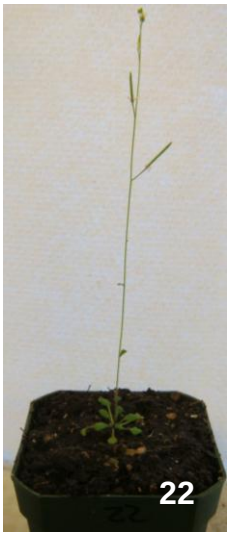


Figure 3.8. (cont.)

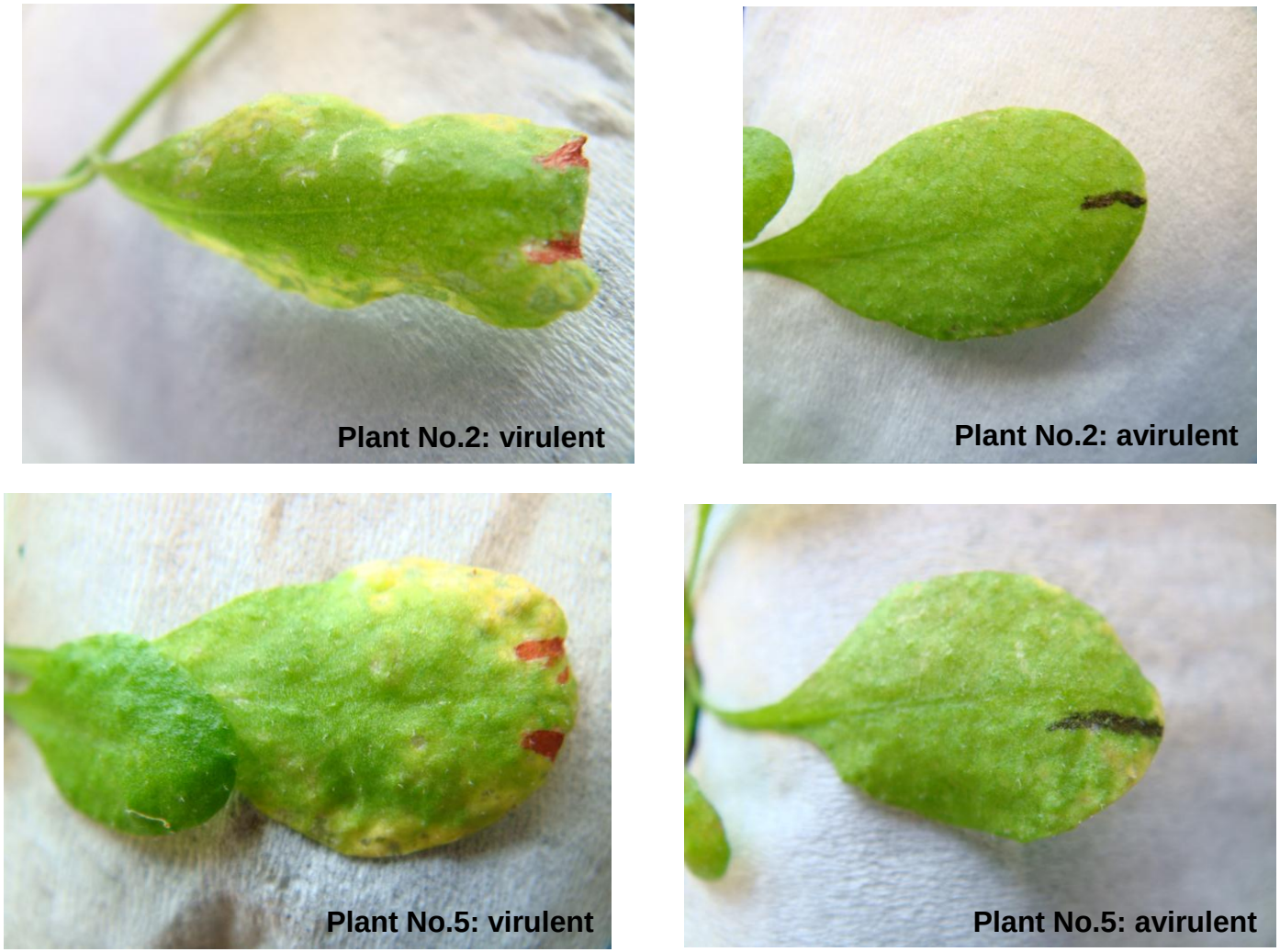


Figure 3.9. Leaves of transgenic plants of DFR/CCR (Glyma18g45260.1) after 48h inoculation. The target leaves were marked with one black dash to indicate inoculated with DC3000 carrying *avrRPT2* and two red markers to indicate inoculated with DC3000 without *avrRPT2*. Pictures were taken at 48hpi before tissue harvest.

Figure 3.9. (cont.)

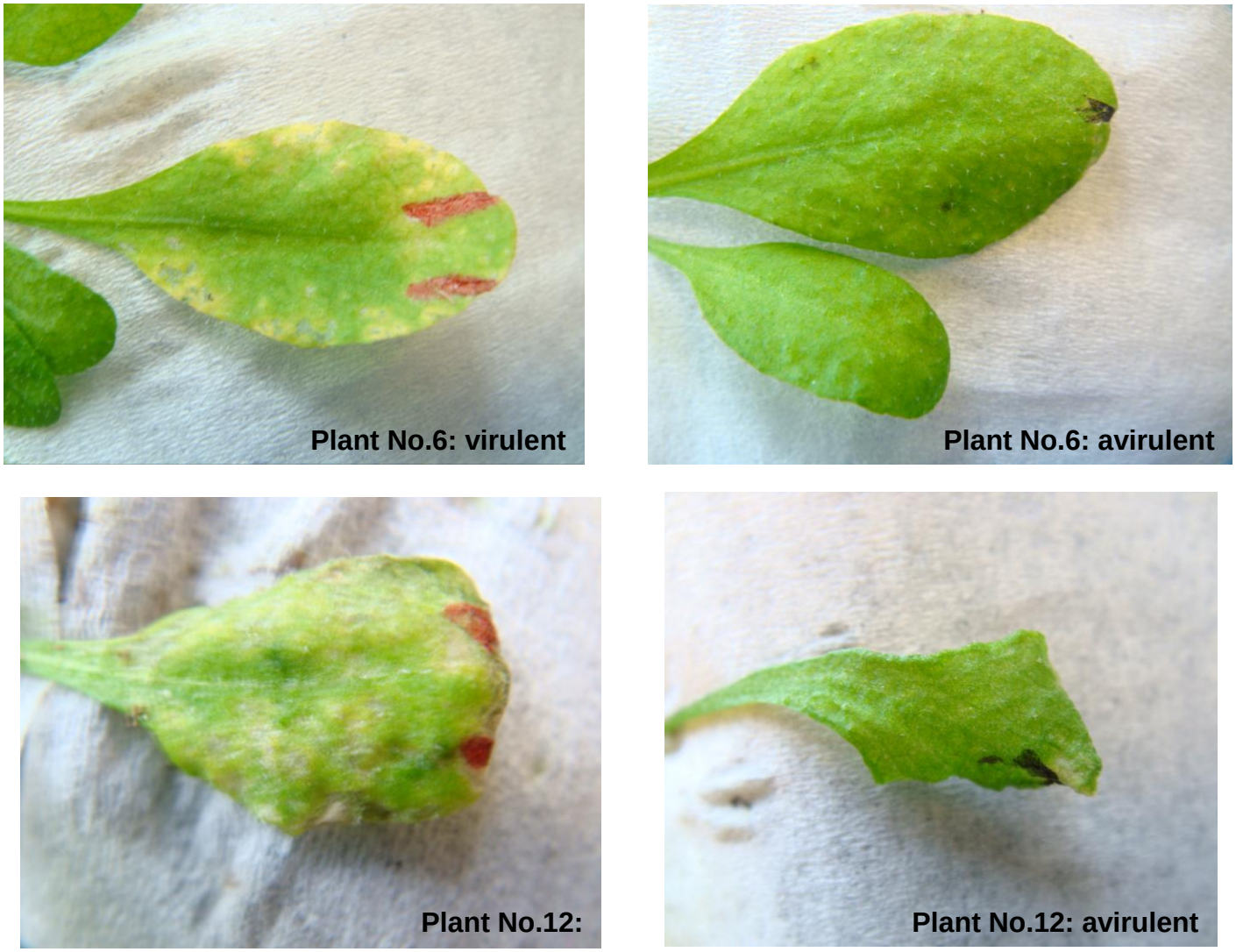


Figure 3.9. (cont.)

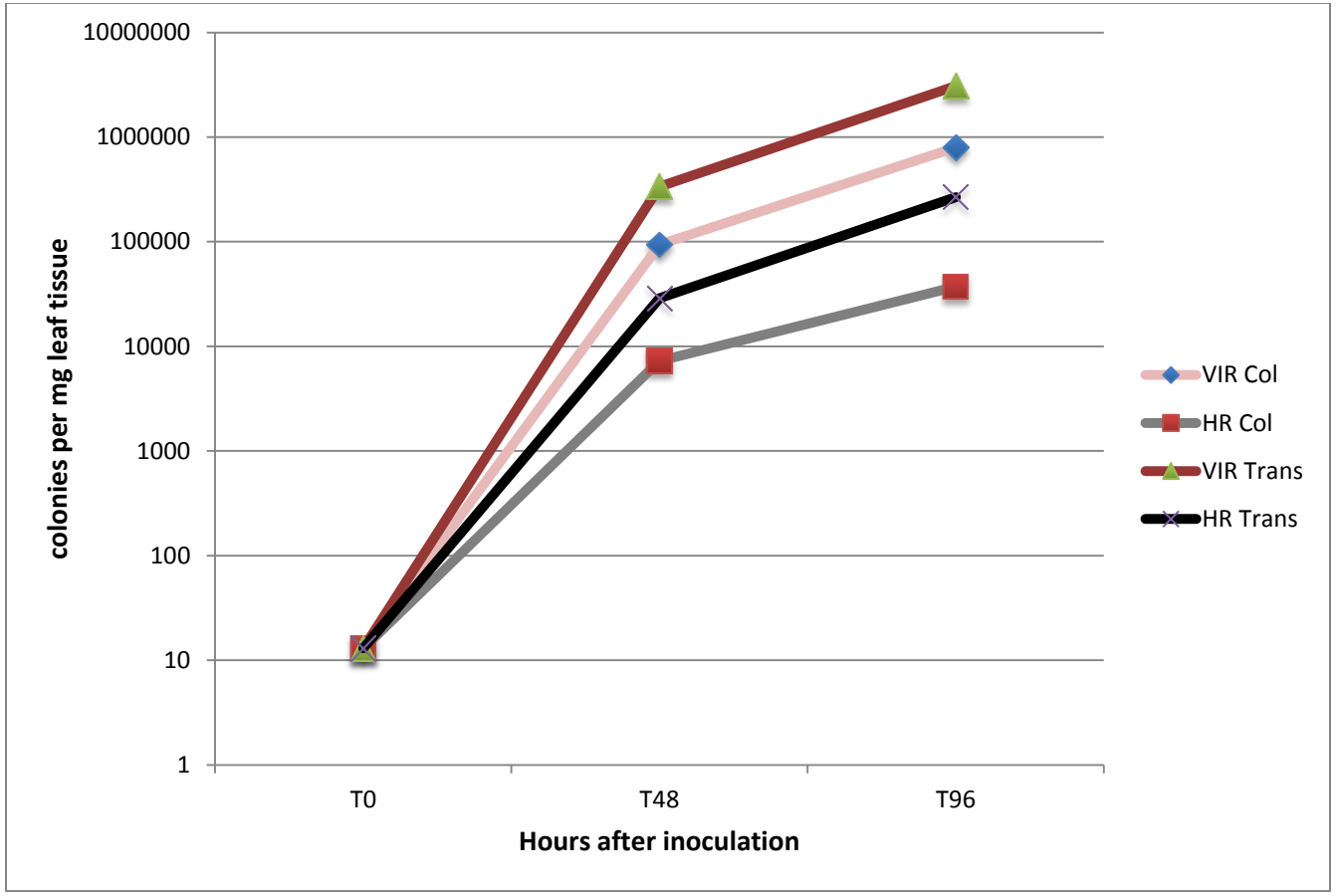


Figure 3.10. Disease assays of *P. syringae* on transgenic Arabidopsis of DFR/CCR (Glyma18g45260.1). *P. syringae* DC3000 virulent (vector control) or avirulent (carrying *avrRpt2*) strains were inoculated into transgenic Arabidopsis leaves. The inoculum contained approximately 5×10^4 cells per ml. Inoculated leaves were collected at 0, 48 and 96 hpi. Resuspended bacteria were plated in serial dilutions. Colonies were counted and used to determine changes in population levels.

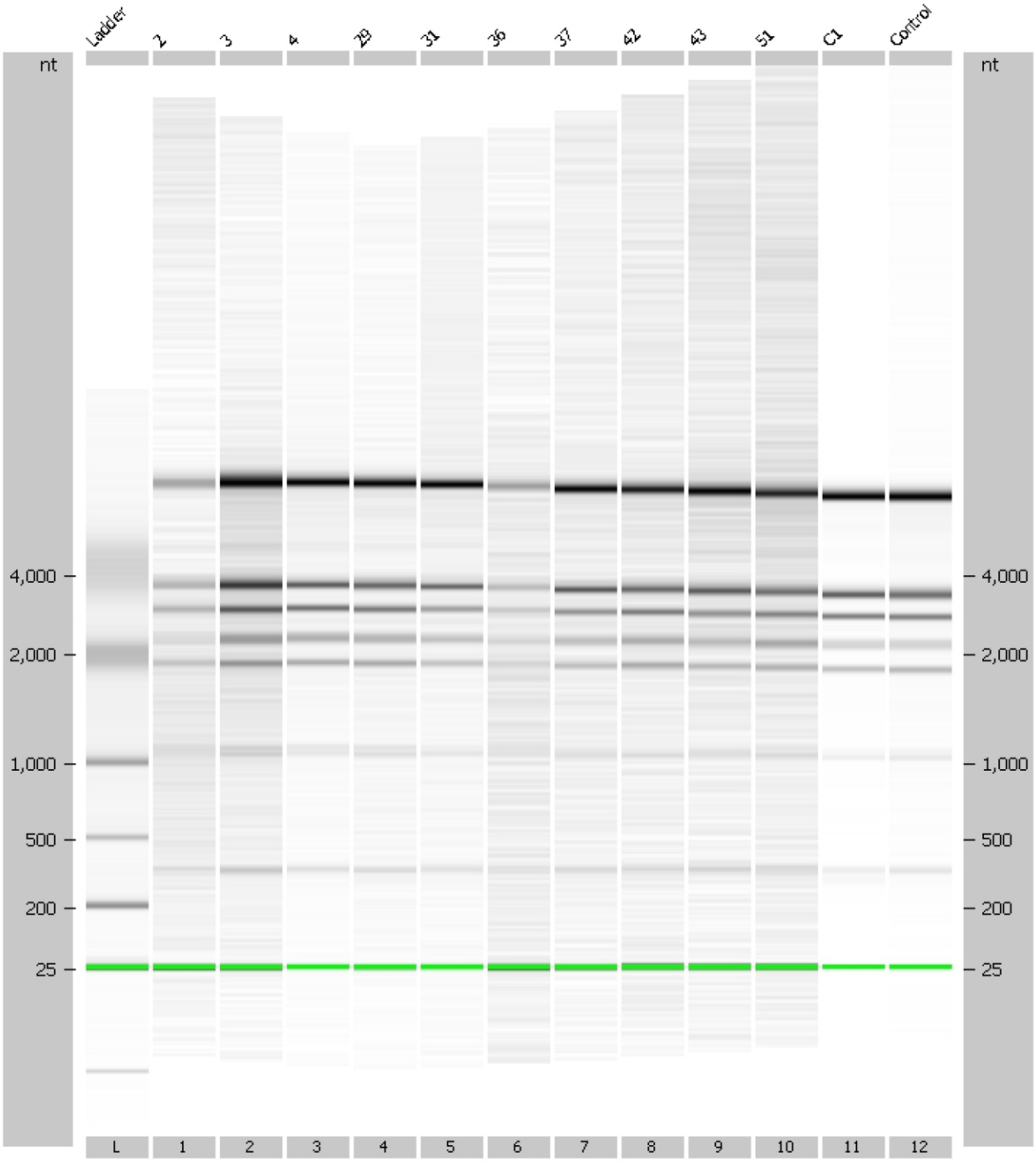
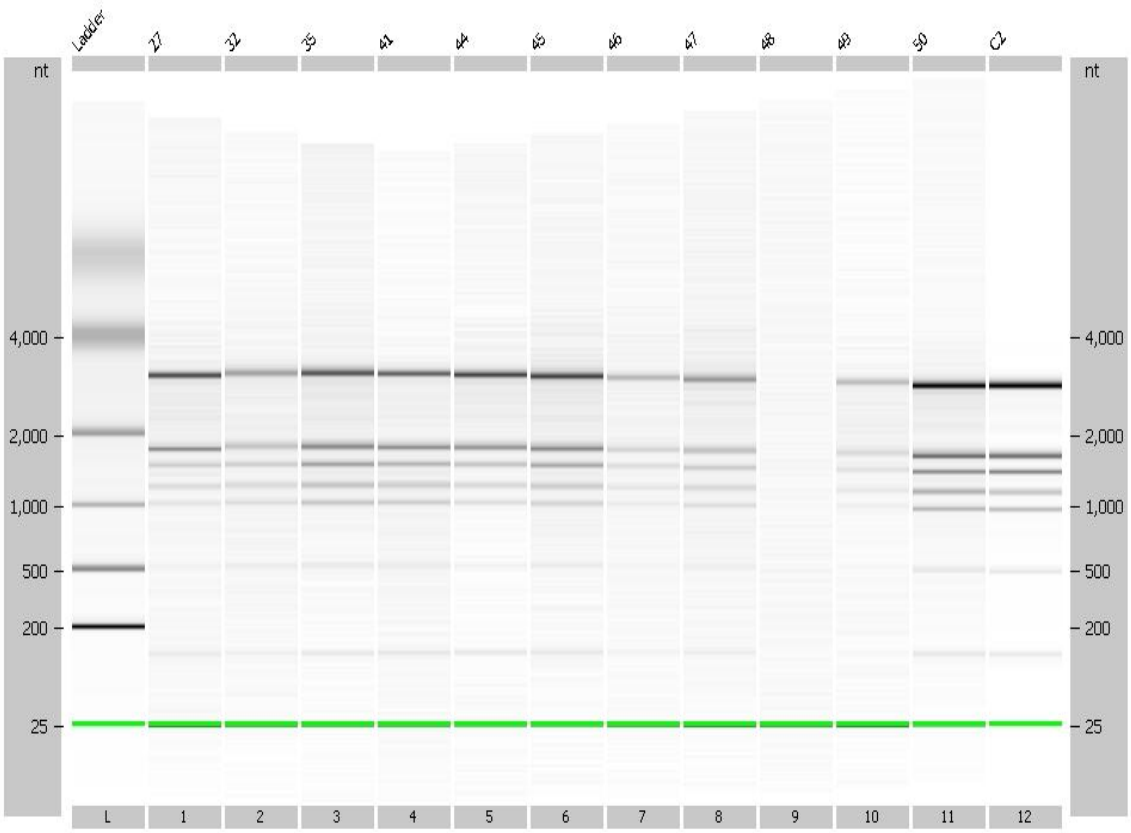
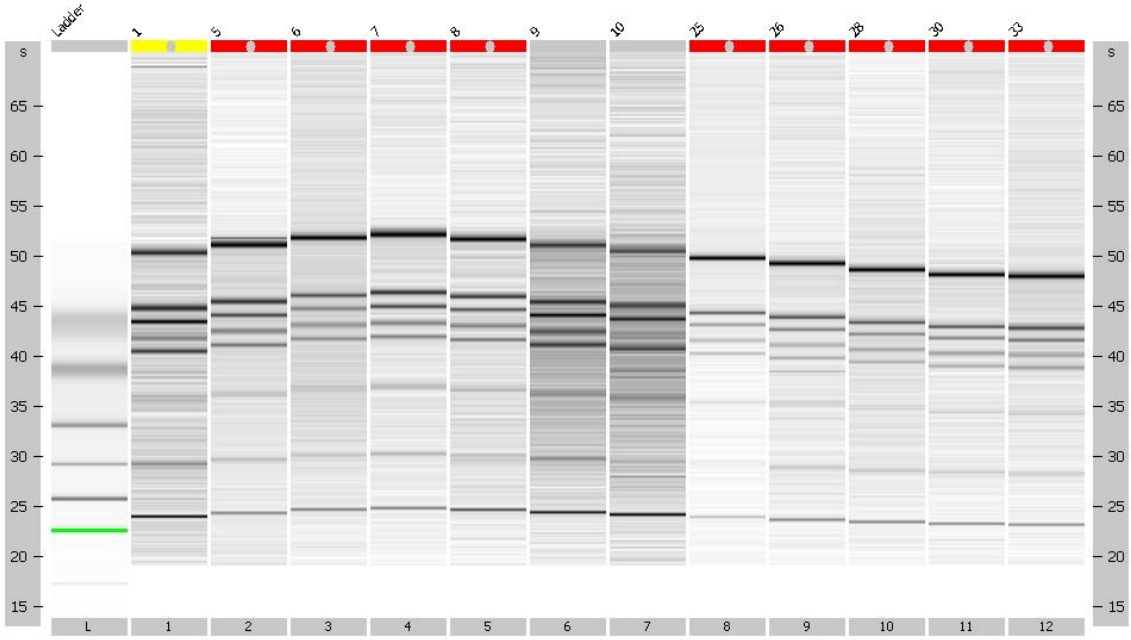


Figure 3.11. Gel result of RNA. RNA quality was determined and confirmed to be high as indicated by the presence the ribosomal RNA bands.

Figure 3.11. (cont.)





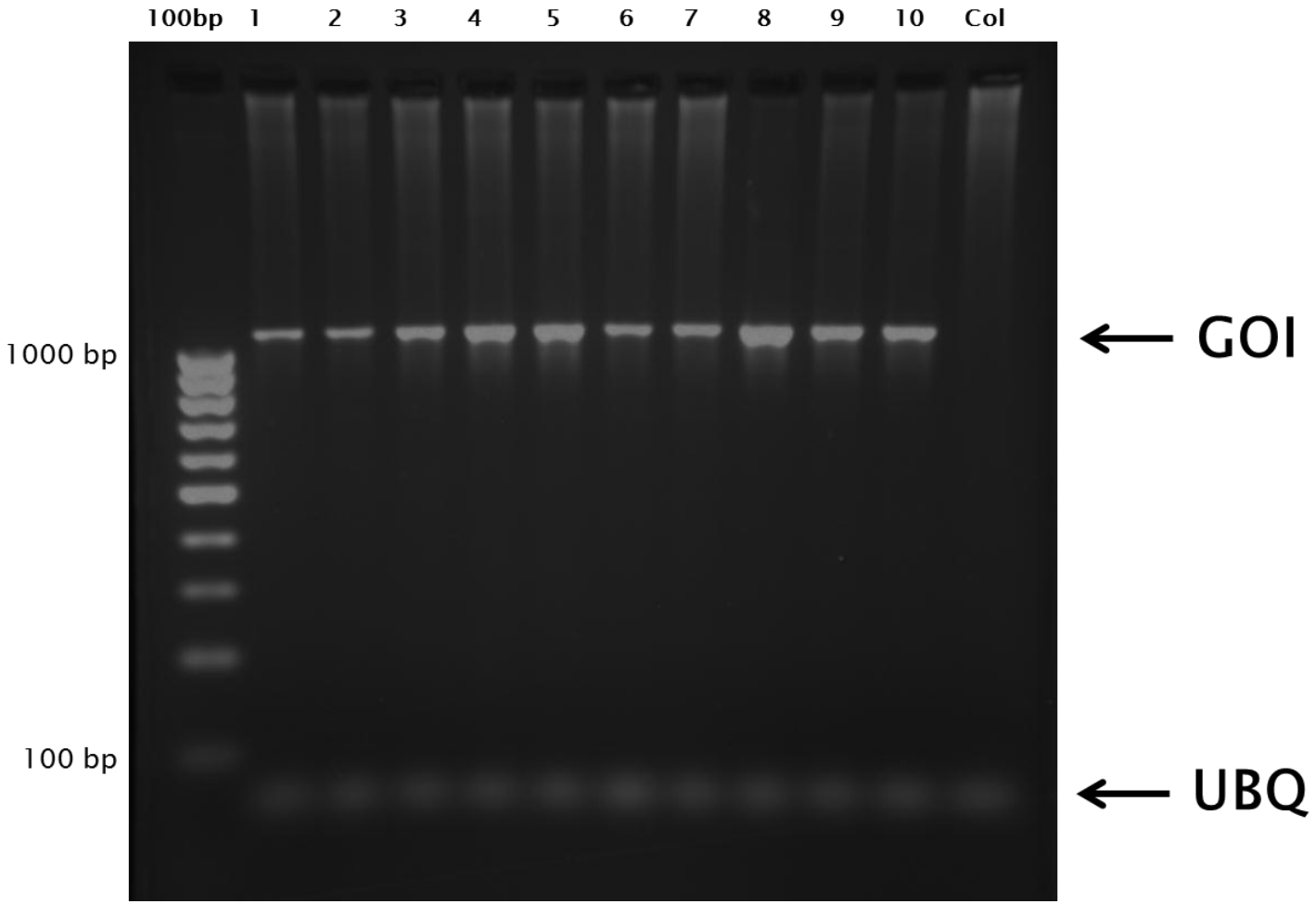


Figure 3.12. Semi-quantitative RT-PCR was performed on the T1 DFR/CCR (Glyma18g45260.1) transgenics. Based on equal loading of RNA and ubiquitin standard control, different levels of expression in each plant were observed. 100bp ladder was used.

CHAPTER IV

CONCLUDING REMARKS

The ability to detect and defend against pathogen infection is an important component of crop productivity. Accordingly, identifying genes involved in pathogen defense is a necessary step to identify pathways for plant genetic improvement.

Our in-house, publicly accessible Soybean Gene Expression Database (SGED) provides a simple, flexible, and powerful gene expression analysis tool. This Master's Thesis presented a study that mined SGED's high-throughput gene expression data across many disease studies for expression patterns of 5,417 genes under 54 treatments, including soybean response to the pathogens *Sclerotinia sclerotiorum* (causal agent of white mold disease), *Fusarium virguliforme* (causal agent of Sudden Death Syndrome disease), and *Pseudomonas syringae* (the causal agent of Bacterial Blight disease). In addition, the expression patterns of these 5,417 genes included their transcription responses to various non-pathogenic treatments (chemicals, Rhizobium, root growth, hairy root growth, and vacuum infiltration stress) to help select pathogen-specific responses. Hierarchical clustering and fuzzy k-means clustering of the data, followed by selective filtering for patterns of interest, allowed for the identification of 11 candidate pathogen-specific responsive genes.

Several of these 11 genes were of particular interest. As transcription responses require transcription factors, it was ensuring to see that a WRKY factor (Glyma11g29720.1) was one of the 11 genes selected, as WRKY factors are often associated with biotic stress responses. A dirigent (Glyma01g31750.1) was also found; these genes are often induced by pathogens (Zou et al., 2005) and are involved in the synthesis of various defensive compounds (Davin and

Lewis, 2000). Two genes were found located next to each other (Glyma17g03340.1 and Glyma17g03350.2), which encode for PR10, a pathogenicity-related (PR) protein, another promising defense-related gene. A primary response to some pathogens is an oxidative burst that serves to prime and induce other defenses, and one of the pathogen-specific genes identified, Glyma19g42220.1, encodes for a possible component of the oxidative burst response. Another, Glyma10g42820.3, encodes for a possible component of the phospholipid metabolism, involved in pathogen signaling. Movement of nutrients and defense metabolites during infection require the presence and function of numerous transport proteins. Two such transporters, an ATP Binding Cassette (ABC) transporter (Glyma10g02370.2) (possible toxin movement) and Glyma18g43540.1 (ammonium transporter) were also found. Three genes were found likely to be involved in secondary metabolism, crucial to plant defense: Glyma18g45260.1 belongs to the phenylpropanoid pathway, often induced in soybean in response to pathogens (Zou et al., 2005; Calla et al., 2009; Radwan et al., 2012). The other two genes (Glyma14g00870.1 and Glyma07g05480.1) encode for enzymes that could alter a substrate to detoxify or to perhaps lead it to be an antimicrobial.

Attempts were made to clone all 11 pathogen induced genes, however, as PCR is tricky in soybean (an ancient tetraploid full of duplications), only six of the 11 were successfully cloned into *E. coli* using the PCR cloning vector pDrive and sequence-verified. Of these six cloned genes, two were successfully cloned into the Agrobacterium binary vector pBin-mgfp5-ER by swapping the *mgfp5* gene with the cDNA of interest using restriction digestion and ligation. The two genes, an O-methyltransferase (OMT) and a dihydroflavonol-4-

reductase/cinnamoyl-CoA reductase protein (DFR/CCR), were successfully transformed into *Arabidopsis thaliana* by floral dip and selection on kanamycin plates.

Disease assays were conducted on the Arabidopsis transgenics to determine if the cloned soybean gene had cross-species effects on plant defense. As the OMT transgenic T1 plants were mostly severely dwarfed, the disease assays for those plants will need to wait until the T2 generation. However, the T1 DFR/CCR transgenics mostly looked healthy and indistinguishable from the parent Col-0 plants, and therefore two individual leaves were inoculated per plant, one leaf with avirulent/incompatible *P. syringae*, and another with virulent/compatible *P. syringae* and bacterial population growth determined over a four day period. Unfortunately, transgenic Arabidopsis plants expressing this *DFR/CCR* gene from soybean did not show a positive enhancement of defense, and if anything, showed a slight increase in susceptibility. These experiments will need to be repeated to see if the results are consistent in the T2 generation.

In the near future, continual attempts will be made to clone and transform the additional genes from this list of 11 pathogen inducible genes into Arabidopsis and eventually soybean. If an effect on disease resistance in soybean is found, their sequences can be developed into molecular markers, allowing breeders to develop more resistant varieties or even develop transgenics with improved resistance. Additionally, this ongoing research project provides genes for promoter analysis and identifies genes responding specifically to pathogen infection, promoters that will be valuable for precision expression of genes of interest in the crops of tomorrow.

REFERENCES

Alkharouf, N. W., Klink, V. P., Chouikha, I. B., Beard, H. S., MacDonald, M. H., Meyer, S., Knap, H.T., & Matthews, B. F. (2006). Timecourse microarray analyses reveal global changes in gene expression of susceptible *Glycine max* (soybean) roots during infection by *Heterodera glycines* (soybean cyst nematode). *Planta*,*224* (4), 838-852.

An, G., Watson, B. D., & Chiang, C. C. (1986). Transformation of tobacco, tomato, potato, and *Arabidopsis thaliana* using a binary Ti vector system. *Plant physiology*, *81*(1), 301-305.

Bent, A. F., Kunkel, B. N., Dahlbeck, D., Brown, K. L., Schmidt, R., Giraudat, J., Leung, J., & Staskawicz, B. J. (1994). RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant disease resistance genes. *Science (New York, NY), 265*(5180), 1856.

Bilgin, D. D., Aldea, M., O'Neill, B. F., Benitez, M., Li, M., Clough, S. J., & DeLucia, E. H. (2008). Elevated ozone alters soybean-virus interaction. *Molecular Plant-Microbe Interactions*, *21*(10), 1297-1308.

Brechenmacher, L., Kim, M. Y., Benitez, M., Li, M., Joshi, T., Calla, B., Lee, M., Libault, M., Vodkin, L., Xu, D., Lee, S., Clough, S., & Stacey, G. (2008). Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. *Molecular plant-microbe interactions*, *21*(5), 631-645.

Cacas, J. L., Petitot, A. S., Bernier, L., Estevan, J., Conejero, G., Mongrand, S., & Fernandez, D. (2011). Identification and characterization of the Non-race specific Disease Resistance 1 (NDR1) orthologous protein in coffee. *BMC plant biology, 11*(1), 144.

Calla, B., Vuong, T., Radwan, O., Hartman, G. L., & Clough, S. J. (2009). Gene expression profiling soybean stem tissue early response to *Sclerotinia sclerotiorum* and in silico mapping in relation to resistance markers. *The Plant Genome*, *2*(2), 149-166.

Cho, S. M., Kang, E. Y., Min, K. H., Lee, Y. K., Kim, Y. C., Yang, K. Y., ... & Cho, B. H. (2012). A positive regulatory role of the watermelon *ClWRKY70* gene for disease resistance in transgenic Arabidopsis thaliana. *Biologia Plantarum, 56*(3), 560-565.

Choi, J. J., Alkharouf, N. W., Schneider, K. T., Matthews, B. F., & Frederick, R. D. (2008). Expression patterns in soybean resistant to *Phakopsora pachyrhizi* reveal the importance of peroxidases and lipoxygenases. *Functional & integrative genomics, 8*(4), 341-359.

Christensen, A. H., Sharrock, R. A., & Quail, P. H. (1992). Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant molecular biology, 18*(4), 675-689.

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant Journal, 16*(6), 735-743.

Davin, L. B., & Lewis, N. G. (2000). Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiology*, *123*(2), 453-462.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, *95*(25), 14863-14868.

Erwin, D. C., & Ribeiro, O. K. (1996). *Phytophthora* diseases worldwide. American Phytopathological Society (APS Press).

Futschik, M. E., & Kasabov, N. K. (2002). Fuzzy clustering of gene expression data. *In Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on (Vol. 1, pp. 414-419). IEEE.*

Gasch, A. P., & Eisen, M. B. (2002). Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol*, *3*(11), 1-22.

Graham, M., Marek, L., & Shoemaker, R. (2002). PCR sampling of disease resistance-like sequences from a disease resistance gene cluster in soybean. *TAG Theoretical and Applied Genetics, 105*(1), 50-57.

Hartman, G. L., Wang, T. C., & Tschanz, A. T. (1991). Soybean rust development and the quantitative relationship between rust severity and soybean yield. *Plant Disease*, *75*(6), 596.

Haseloff, J., Siemering, K. R., Prasher, D. C., & Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proceedings of the National Academy of Sciences*, *94*(6), 2122-2127.

Hinsch, M., & Staskawicz, B. (1996). Identification of a new Arabidopsis disease resistance locus, RPS4, and cloning of the corresponding avirulence gene, *avrRps4*, from *Pseudomonas syringae* pv. *pisi*. *MPMI-Molecular Plant Microbe Interactions*, 9(1), 55-61.

Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J., & Schilperoort, R. A. (1983). A binary plant vector strategy based on separation of vir-and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature*, 179-180.

Kachroo, A., & Kachroo, P. (2009). Fatty acid-derived signals in plant defense. *Annual review of phytopathology*, *47*, 153-176.

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R. J., Bilgin, D. D., Radwan, O., Neece, D. J., Clough, S. J., May, G. D., & Stacey, G. (2010). Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant physiology, 152*(2), 541-552.

Moy, P., Qutob, D., Chapman, B. P., Atkinson, I., & Gijzen, M. (2004). Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. *Molecular Plant-Microbe Interactions*, *17*(10), 1051-1062.

Panthee, D. R., Yuan, J. S., Wright, D. L., Marois, J. J., Mailhot, D., & Stewart, C. N. (2007). Gene expression analysis in soybean in response to the causal agent of Asian soybean rust (*Phakopsora pachyrhizi Sydow*) in an early growth stage. *Functional & integrative genomics*, 7(4), 291-301.

Quackenbush, J. (2001). Computational analysis of microarray data. *Nature Reviews Genetics*, *2*(6), 418-427.

Quirino, B. F., Normanly, J., & Amasino, R. M. (1999). Diverse range of gene activity during *Arabidopsis thaliana* leaf senescence includes pathogen-independent induction of defense-related genes. *Plant molecular biology*, *40*(2), 267-278.

Radwan, O., Liu, Y., & Clough, S. J. (2011). Transcriptional analysis of soybean root response to *Fusarium virguliforme*, the causal agent of sudden death syndrome. *Molecular Plant-Microbe Interactions*, *24*(8), 958-972.

Rolfe, S. A., & Tobin, E. M. (1991). Deletion analysis of a phytochrome-regulated monocot rbcS promoter in a transient assay system. *Proceedings of the National Academy of Sciences, 88*(7), 2683-2686.

Rushton, P. J., & Somssich, I. E. (1998). Transcriptional control of plant genes responsive to pathogens. *Current opinion in plant biology*, *1*(4), 311-315.

Scheideler, M., Schlaich, N. L., Fellenberg, K., Beissbarth, T., Hauser, N. C., Vingron, M., Slusarenko, A.J., & Hoheisel, J. D. (2002). Monitoring the switch from housekeeping to pathogen defense metabolism in *Arabidopsis thaliana* using cDNA arrays. *Journal of Biological Chemistry*, *277*(12), 10555-10561.

Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H. S., Han, B., Zou, G., & Katagiri, F. (2003). Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *The Plant Cell Online*, *15*(2), 317-330.

Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., & Church, G. M. (1999). Systematic determination of genetic network architecture. *Nature genetics*, *22*, 281-285.

Torres, M. A., & Dangl, J. L. (2005). Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. *Current opinion in plant biology*, *8*(4), 397-403.

Ülker, B., & Somssich, I. E. (2004). WRKY transcription factors: from DNA binding towards biological function. *Current opinion in plant biology*, *7*(5), 491-498.

Vaghchhipawala, Z., Bassüner, R., Clayton, K., Lewers, K., Shoemaker, R., & Mackenzie, S. (2001). Modulations in gene expression and mapping of genes associated with cyst nematode infection of soybean. *Molecular plant-microbe interactions, 14*(1), 42-54.

van de Mortel, M., Recknor, J. C., Graham, M. A., Nettleton, D., Dittman, J. D., Nelson, R. T., Godoy, C. V., Abdelnoor, R. V., Almeida, A. M. R., Baum, T. J., & Whitham, S. A. (2007). Distinct biphasic mRNA changes in response to Asian soybean rust infection. *Molecular plant-microbe interactions, 20*(8), 887-899.

Van Loon, L. C., Rep, M., & Pieterse, C. M. J. (2006). Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.*, *44*, 135-162.

Vodkin, L. O., Khanna, A., Shealy, R., Clough, S. J., Gonzalez, D. O., Philip, R., Thibaud-Nissen, F., Sidarous, M., Stromvik, M. V., Shoop, E. G., Schmidt, C., Retzel, E., Erpelding, J., Shoemaker, R. C., Rodrizues-Huete, A. M., Polacco, J. C., Coryell, V., Keim, P., Gong, G., Lui, L., Pardinas, J. L., & Schweitzer, P. (2004). Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. *BMC genomics*, *5*(1), 73.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57-63.

Wei, W., Huang, J., Hao, Y. J., Zou, H. F., Wang, H. W., Zhao, J. Y., Liu, X. Y., Zhang, W. K., Ma, B., Zhang, J. S., & Chen, S. Y. (2009). Soybean GmPHD-type transcription regulators improve stress tolerance in transgenic Arabidopsis plants. *PLoS One, 4*(9), e7209.

Whalen, M. C., Innes, R. W., Bent, A. F., & Staskawicz, B. J. (1991). Identification of *Pseudomonas syringae* pathogens of Arabidopsis and a bacterial locus determining avirulence on both Arabidopsis and soybean. *The Plant Cell Online*, *3*(1), 49-59.

Yang, H., Huang, Y., Zhi, H., & Yu, D. (2011). Proteomics-based analysis of novel genes involved in response toward soybean mosaic virus infection. *Molecular biology reports, 38*(1), 511-521.

Zhu, J., Patzoldt, W. L., Radwan, O., Tranel, P. J., & Clough, S. J. (2009). Effects of photosystem-II-interfering herbicides atrazine and bentazon on the soybean transcriptome. *The Plant Genome*, *2*(2), 191-205.

Zhu, J., Patzoldt, W. L., Shealy, R. T., Vodkin, L. O., Clough, S. J., & Tranel, P. J. (2008). Transcriptome response to glyphosate in sensitive and resistant soybean. *Journal of agricultural and food chemistry*, *56*(15), 6355-6363.

Zou, J., Rodriguez-Zas, S., Aldea, M., Li, M., Zhu, J., Gonzalez, D. O., Vodkin, L., DeLucia, E., & Clough, S. J. (2005). Expression profiling soybean response to *Pseudomonas syringae* reveals new defense-related genes and rapid HR-specific downregulation of photosynthesis. *Molecular plant-microbe interactions*, *18*(11), 1161-1174.